

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 24, 2003, 19:13:28 ; Search time 62 Seconds
(without alignments)
5721.180 Million cell updates/sec

Title: US-09-830-244b-2
Perfect score: 2322
Sequence: 1 ccatgttttttgatcaaaag.....gcctttccatgctcactg 1331

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 908470 segs, 133250620 residues
Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ .n2p .model -DEV=xlp
-O=cg2_1/USPRO.spool/US09830244/runat.24062003.130000.10004/app_query.fasta.1.1479
-DB=geneseq.101002 -OEMT=fastcan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCT=0
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonuma2 -TRANS=human40.cdl
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MNTLEN=0 -MAXLEN=200000000
-USRR=US09830244 -GCCH_1.1.114 -struna_24062003.130000.10004 -NCPU=6 -ICPU=3
-NO_MAMP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MAMP_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq.101002.*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	522	22.5	95	21	AAV94526	Human lysine-rich
2	136	5.9	62	21	AAV94527	Human statherin pr
3	136	5.9	82	21	AAV42456	Human. ORF ORF2220
4	93	4.0	384	20	AAV00147	Enterococcus faeca
5	93	4.0	384	20	ABP43366	E faecalis EF072 a
6	93	4.0	430	20	AAV00146	Enterococcus faeca
7	87.5	3.7	360	16	AAV00145	Enterococcus faeca
8	85.5	3.7	360	16	AAV00144	Enterococcus faeca
9	85.5	3.7	360	16	AAV00143	Enterococcus faeca
10	85.5	3.7	360	16	AAV00142	Enterococcus faeca
11	85.5	3.7	360	16	AAV00141	Enterococcus faeca
12	85.5	3.7	360	16	AAV00140	Enterococcus faeca
13	85.5	3.7	360	16	AAV00139	Enterococcus faeca
14	85.5	3.7	360	16	AAV00138	Enterococcus faeca
15	85	3.7	360	16	AAV00137	Enterococcus faeca
16	85	3.7	360	16	AAV00136	Enterococcus faeca
17	84.5	3.7	360	16	AAV00135	Enterococcus faeca
18	84.5	3.7	360	16	AAV00134	Enterococcus faeca
19	84.5	3.7	360	16	AAV00133	Enterococcus faeca
20	84.5	3.7	360	16	AAV00132	Enterococcus faeca
21	84.5	3.7	360	16	AAV00131	Enterococcus faeca
22	84.5	3.7	360	16	AAV00130	Enterococcus faeca
23	84.5	3.7	360	16	AAV00129	Enterococcus faeca
24	84	3.6	378	20	ABP4255	Human IL-8 recepto
25	84	3.6	378	20	ABP4254	Human IL-8 recepto
26	84	3.6	378	20	ABP4253	Human IL-8 recepto
27	84	3.6	378	20	ABP4252	Human IL-8 recepto
28	84	3.6	378	20	ABP4251	Human IL-8 recepto
29	84	3.6	378	20	ABP4250	Human IL-8 recepto
30	84	3.6	378	20	ABP4249	Human IL-8 recepto
31	83.5	3.6	378	20	ABP4248	Human IL-8 recepto
32	83.5	3.6	378	20	ABP4247	Human IL-8 recepto
33	83.5	3.6	378	20	ABP4246	Human IL-8 recepto
34	83.5	3.6	378	20	ABP4245	Human IL-8 recepto
35	82.5	3.6	378	20	ABP4244	Human IL-8 recepto
36	82.5	3.6	378	20	ABP4243	Human IL-8 recepto
37	82.5	3.6	378	20	ABP4242	Human IL-8 recepto
38	82.5	3.6	378	20	ABP4241	Human IL-8 recepto
39	82.5	3.6	378	20	ABP4240	Human IL-8 recepto
40	82	3.5	378	20	ABP4239	Human IL-8 recepto
41	82	3.5	378	20	ABP4238	Human IL-8 recepto
42	82	3.5	378	20	ABP4237	Human IL-8 recepto
43	81.5	3.5	378	20	ABP4236	Human IL-8 recepto
44	81.5	3.5	378	20	ABP4235	Human IL-8 recepto
45	81	3.5	378	20	ABP4234	Human IL-8 recepto

ALIGNMENTS

RESULT 1
ID: AAV94526 standard: protein: 95 AA.
AAV94526;
06-OCT-2000 (first entry)

Human Lysine-rich statherin protein.

Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human; precipitation inhibitor; autoimmune; inflammatory disorder; AIDS; asthma; allergy; diabetes mellitus; fungal; bacterial infection; cancer; leukemia; adenocarcinoma; melanoma.

Key: Modified-site 15 Location/Qualifiers

1

Alignment Scores:

Pred. No.:	0.0409	Length:	384
Score:	93.00	Matches:	74
Percent Similarity:	36.26%	Conservative:	21
Best Local Similarity:	28.24%	Mismatches:	85
Query Match:	4.02%	Indels:	83
DB:	20	Gaps:	12

US-09-830-244B-2 (1-1331) x AAY00147 (1-384)

```

QY 672 ACTGTCCTGCTCCAGAGGGATTCATGAGATGCTAAGAGGAAACACCTGCAG 613
DB 149 ThrValLeuAspLysGlnGlnGlyPheAsn-----Gln 159
QY 612 GGGAGCAGCATCTTGATTCAGATTAAG-----TAC 580.
DB 160 Gly-----GlnHisIleAsnTyrGlnLeuThrGlnIleProAlaAsnIleuGlyTyr 178
QY 579 CAAAAAATATTAGCTATCTCTTTCGACACATGCGCTCTTCTGTTACACTTTCAGT 520
DB 179 Gln-GluPheArgLeuSerAspLysAlaAspThrThrLeuThrLeuProGluSerIle 198
QY 519 CATTCAAAAGCGCTATTGCTAGAG-----TGCAGCAATCCAGTG 479
DB 198 egluValIysValAlaGlyLysThrValThrThrGlyTyrThrLeuThrGlnLysHis 218
QY 478 TAGATTACTGTAACAGTTATTATTA-----ACTATGCT 446
DB 218 sglYpHeThrLeuAspPheSerIleLysAspLeuGlnAsnPhelaAsnGlnThrMetTh 238
QY 445 GTATTCA-----CCAATAAGCGCATTCACCAATAATGCA----- 412
DB 238 rValSerTyrGlnMetArgLeuGlnLysThrAlaGluProAspThrAlaIleAsnAsnGln 258
QY 411 -----TTAGGCATCAAAAGATCTGAAGATGCTAATGGACATTCGCCACAAAAG 359
DB 258 uglYglnLeuValThrAspLysHisThrLeuThrLysArgAlaThrValArgThrGlyGln 278
QY 358 AAGC-----AAATCCATGCGCTCTATA----- 336
DB 278 YLysSerPheValIysValAspSerGluAsn-AlaLysIleThrLeuProGluAlaValP 298
QY 335 -----AAGGAGATATATAGCAATGGCAGACAGAAATGAAATTTCTGTC 293
DB 298 heIleValIysAsnGlnAlaGlyLysThrLeuAsnGlnThrAlaAsnGlyTyrArgTyrpg 318
QY 292 AAAAGAAAGTACTTCTCAAGCAATCAGAAACCTGTCCCAATTAAGCGCTGGAGC 233
DB 318 LnyLysGlu-----LysAlaLeuAlaLysLysPheThrSerAsnGlnAlaGlyLys 334
QY 232 CACATTAGCCACTATCTCAAGTACTGATATTGTTTACCTTACGGGCGCTCAATATTC 173
DB 335 -----PheSerValLysGly**LysArgT 343
QY 172 CCCCTCATCTGCTAAACTCACTGTTTACTCACAAGTGTGATTAATTCACAGCAAGTA 113
DB 343 rPrroValIleuLeuGlnArgAsnLeuGlyThrLysArgLeuSer-----SerGlnSerA 361
QY 112 ACCTTAACATTCCTCATGCGAGCTGCAAAATTCAGACATTTGTAACACTGTAATTTCA 53
DB 361 snArgAsnSerPheTyrGlyGlyLysLysPheLeuGlyAsnGlnLysThrThrAsnSer 381
QY 52 CG 51
DB 381 hr 381

```

RESULT 5
ABP43366 standard; Protein; 384 AA.

ID ABP43366;
AC ABP43366;
XX
XX

05-AUG-2002 (first entry)

E faecalis EF072 antigenic fragment.

Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic.

Enterococcus faecalis.

US2002045737-A1.

18-APR-2002.

04-MAY-1998; 98US-0071035.

04-MAY-1998; 98US-0071035.

(HUMA-) HUMAN GENOME SCI INC.

Choi GH, Bailey C, Hromockyj A, Kunsch CA;

WPI; 2002-425450/45.

N-PSDB; ABW98122.

New genes and polypeptides from *Enterococcus faecalis*, useful as vaccines for preventing, treating or attenuating an infection caused by a member of the *Enterococcus* genus in an animal, particularly *E. faecalis*.

Claim 9; Page 133; 255pp; English.

The present invention provides the protein and coding sequences of a number of polypeptides from *Enterococcus faecalis*. The proteins can be used as vaccines for preventing or attenuating an infection caused by a member of the *Enterococcus* genus in an animal, particularly *E. faecalis*. The polynucleotide is also useful for preventing or treating *E. faecalis* infection. The present sequence is a protein of the invention.

Sequence 384 AA;

Alignment Scores:

Pred. No.:	0.0409	Length:	384
Score:	93.00	Matches:	74
Percent Similarity:	36.26%	Conservative:	21
Best Local Similarity:	28.24%	Mismatches:	85
Query Match:	4.02%	Indels:	83
DB:	23	Gaps:	12

US-09-830-244B-2 (1-1331) x ABP43366 (1-384)

```

QY 672 ACTGTCCTGCTCCAGAGGGATTCATGAGATGCTAAGAGGAAACACCTGCAG 613
DB 149 ThrValLeuAspLysGlnGlnGlyPheAsn-----Gln 159
QY 612 GGGAGCAGCATCTTGATTCAGATTAAG-----TAC 580
DB 160 Gly-----GlnHisIleAsnTyrGlnLeuThrGlnIleProAlaAsnIleuGlyTyr 178
QY 579 CAAAAAATATTAGCTATCTCTTTCGACACATGCGCTCTTCTGTTACACTTTCAGT 520
DB 179 Gln-GluPheArgLeuSerAspLysAlaAspThrThrLeuThrLeuProGluSerIle 198
QY 519 CATTCAAAAGCGCTATTGCTAGAG-----TGCAGCAATCCAGTG 479
DB 198 egluValIysValAlaGlyLysThrValThrThrGlyTyrThrLeuThrGlnLysHis 218
QY 478 TAGATTACTGTAACAGTTATTATTA-----ACTATGCT 446
DB 218 sglYpHeThrLeuAspPheSerIleLysAspLeuGlnAsnPhelaAsnGlnThrMetTh 238
QY 445 GTATTCA-----CCAATAAGCGCATTCACCAATAATGCA----- 412
DB 238 rValSerTyrGlnMetArgLeuGlnLysThrAlaGluProAspThrAlaIleAsnAsnGln 258

```


QY 411 -----TTAGTCACATCAAAAGATCTGAATGCTAATGGACATTTCCACAAAAG 359
 Db 258 uclglnleuvalthrpsrlysthrleuthrlysalathrvalargthglycl 278
 QY 358 AAG-----AAATCCATGCTCTCTATA----- 336
 Db 278 ylyserphevallysalaspsersgluasn-AlaLysllethrleuproglnalvalp 298
 QY 335 -----AAGGAGATATAGCAATGGCAACCAAGATGAATTTCTGTC 293
 Db 298 heilevallysalasnglnalaglyglutyrleuasnglutrhalaasnlytyrargtrpg 318
 QY 292 AAAAGAAAGTCTCTCAACCAATCAAGAAAACCTGTTCCCAATTAAGCCCTGGGAGC 233
 Db 318 lnylsglu-----LysAlaLeuAlaLysLysPheThrSerAsnGlnAlaLysGlu- 334
 QY 232 CACATTAAGCCACATCTCAGATGATTAATTTGTTAGCCTTAAGGGCCTCAATATTC 173
 Db 335 -----PheSerValLysGly**LysArgT 343
 QY 172 CCCCTCATCTGCTAAACTCAACTGTTTACTCAGATGATGATTAATTCACGCAAGTA 113
 Db 343 rprvalleuvalgluargasnleucysphrlysargluSer-----SerGluSerA 361
 QY 112 ACCTTAACATTCCTCATGCGCAGTGCAGAAATTCAGACATTTGTAACACTGTAATTTCA 53
 Db 361 snargasnserpheytyrlyglylylsphleucysasnngluargthtrthasnsert 381
 QY 52 CG 51
 Db 381 hr 381

RESULT 6

AAY00146
 ID AAY00146 standard; Protein; 430 AA.

AC AAY00146;
 XX 20-APR-1999 (first entry)
 DT Enterococcus faecalis protein EF072.
 DE Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 KM detection; attenuation; antigenic.
 XX Enterococcus faecalis.
 OS
 PN MO9850554-A2.
 XX 12-NOV-1998.
 PD 04-MAY-1998; 98WO-US08959.
 XX 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Bailey C, Choi GH, Hromocky A, Kunsch CA;
 DR WPI; 1999-070095/06.
 XX N-PSDB: AAX20136.
 XX
 PT New isolated Enterococcus faecalis polynucleotides - used to develop
 PT products for the detection of Enterococcus and for use in vaccines
 PT for prevention or attenuation of Enterococcus infection
 XX
 PS Claim 9; Page 163; 301pp; English.

CC The present sequence represents a protein isolated from
 CC Enterococcus faecalis. The present invention describes genes, proteins
 CC and antigenic polypeptides isolated from E. faecalis. The proteins can

CC be used in vaccines for preventing or attenuating an infection caused
 CC by a member of the Enterococcus genus in an animal. They can also be
 CC used for detecting Enterococcus antibodies in a sample. The nucleotide
 CC sequences can be used for detecting Enterococcus nucleic acids.
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of E. faecalis protein
 CC activity.

SQ Sequence 430 AA;
 Alignment Scores:
 Pred. No.: 0.0429 Length: 430
 Score: 93.00 Matches: 74
 Percent Similarity: 36.26% Conservative: 21
 Best Local Similarity: 28.24% Mismatches: 85
 Query Match: 4.02% Indels: 83
 DB: 20 Gaps: 12

US-09-830-244B-2 (1-1331) x AAY00146 (1-430)

QY 672 ACTGCTCTGCTCCAGAGGGATTCATGAGATCGTAAGAAAGCAACCTGCAG 613
 Db 193 ThrValLeuAspLysGlnGlnGlyPheAsn-----Gln 203
 QY 612 GGGAGACAGACCTTGATTCAGATTAAG-----TNC 580
 Db 204 Gly---GluHisLysAsnTyrglnleuThrThrGlnleuProAlaAsnleuGlytyr 222
 QY 579 CAAAAATATTAAGCTATCTCTTTCGACACATGCTTTCTGTTAACAATTTCACT 520
 Db 223 Gln-GluPheArgLeuSerAspLysAlaAspThrThrleuThrleuProGlnSeril 242
 QY 519 CATTCAAAAGCCTATTGTAGAG-----TCACGCAATTCACAGTG 479
 Db 242 egluvallysalalaglylysthrvalthrthrglytyrthrleuthrthrGlnLysHl 262
 QY 478 TAGATTACTTCAACAGTTTATTTA-----ACTATGCT 446
 Db 262 sglpheythrleuaspheSerlelelysalpsleuGlnsnPheAlaAsnGlnThrMeth 282
 QY 445 GTATTC-----CCAAATACGGCATTTACCAATATGCA----- 412
 Db 282 rValSerTyrglnMetArgLeuGlnLysThrAlaGluProAspThrAlaLeuAsnGln 302
 QY 411 -----TTAGTCACATCAAAAGATCTGAATAATGCTAATGACATTTCCACAAAAG 359
 Db 302 uclglnleuvalthrpsrlysthrleuthrlysalathrvalargthglycl 322
 QY 358 AAG-----AAATCCATGCTCTCTATA----- 336
 Db 322 ylyserphevallysalaspsersgluasn-AlaLysllethrleuproglnalvalp 342
 QY 335 -----AAGGAGATATAGCAATGGCAACCAAGATGAATTTCTGTC 293
 Db 342 heilevallysalasnglnalaglyglutyrleuasnglutrhalaasnlytyrargtrpg 362
 QY 292 AAAAGAAAGTCTCTCAACCAATCAAGAAAACCTGTTCCCAATTAAGCCCTGGGAGC 233
 Db 362 lnylsglu-----LysAlaLeuAlaLysLysPheThrSerAsnGlnAlaLysGlu- 378
 QY 232 CACATTAAGCCACATCTCAGATGATTAATTTGTTAGCCTTAAGGGCCTCAATATTC 173
 Db 379 -----PheSerValLysGly**LysArgT 387
 QY 172 CCCCTCATCTGCTAAACTCAACTGTTTACTCAGATGATGATTAATTCACGCAAGTA 113
 Db 387 rprvalleuvalgluargasnleucysphrlysargluSer-----SerGluSerA 405
 QY 112 ACCTTAACATTCCTCATGCGCAGTGCAGAAATTCAGACATTTGTAACACTGTAATTTCA 53
 Db 405 snargasnserpheytyrlyglylylsphleucysasnngluargthtrthasnsert 425
 QY 52 CG 51

The invention relates to identifying modulators of receptor activity using orthologues of human receptors by contacting a compound with receptors from two species, and measuring the effect of the compound on the receptors. Included is an isolated nucleic acid (I) comprising a nucleotide sequence encoding bradykinin B1 receptor (II); or encoding a CXCR2 receptor (III). The method is useful for identifying a compound having dual specificity to modulate the activity of a desired polypeptide in two different species. The method is useful for identifying a non-human animal model for testing compounds with potential efficacy as human receptor modulators. The method comprises contacting a test compound with receptors from at least two species; measuring an effect of the compound on the receptors; and selecting an animal model representing a species having a receptor that exhibits the desired effect when contacted with the test compound. Also described is a method for identifying a compound that modulates (II) or (III) activity, where a compound is an agonist, antagonist, suppressor, inhibitor and inducer, useful in treatment of pain and inflammation and other receptor related pathologies. AA080476-AA080492 represent bradykinin I and CXCR2 receptor amino acid sequences of the invention.

```

      :: || || | |||||:::||
NetG]vAsnAntBr.]Aacm--aa- 6/6
-----
195 -----

```

RESULT 9	
AA28273	
ID	AA28273 standard; Protein; 360 AA.
XX	
AC	AA28273;
XX	
DT	04-APR-1993 (first entry)
XX	

screened with a 652 bp *ECORI*/

disclosure; Fig 2; 71pp; English.

Rabbit high affinity IL-8 receptor gene was isolated from rabbit peritoneal neutrophils and used as a source of poly(A)⁺ RNA, to produce a rabbit neutrophil cDNA library. 250,000 recombinant plaques were screened for those which hybridized to an antisense oligonucleotide (AAQ30015). This probe was designed to base on the sequence derived from the second transmembrane domain of G-protein-coupled receptors. After tertiary screening, six plaques were isolated. The insert of one of these plaques, termed F3R was of 2.5 kb in size. This insert was sequenced. The protein deduced from the F3R clone demonstrates that it belongs to the family of G-protein-coupled receptors. The deduced protein sequence indicates seven putative transmembrane segments. A human peripheral blood leukocyte lambda g11 cDNA library (including nucleotides 27 to 625) of the rabbit F3R clone. After tertiary screening several human clones which hybridized to the rabbit IL-8 probe were isolated. The insert of one such clone, termed 4AB was sequenced (AAQ30012).

XX SQ Sequence 360 AA:

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
0.344	85.50	37.84%	22.39%	3.70%	360	58	40	87	75	10

US-09-830-244B-2 (1-1331) x AAR28273 (1-360)

```

YY 1187 AGTCGATTTTGGCCATCAAGCCCTCAATATCATGCTCTGCTACTGATG-----ATA 1134
DB 76 SerArgValGlyArgSerValThrAspValTyrLeuLeuGlnLeuAlaLeuAlaAspLeu 95
YY 1133 TTTAAACGATATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1074
DB 96 LeuPheAlaLeuThrLeuProIleTyrPalaIaSerLysValAsnGlyTyrPheGly 115
YY 1073 AACTGCGTATAGGCTCTTGGTATAAAGCAATATCCAGAAATTAAGATTAATTAATTT 1014
DB 116 ThrPheLeu-----CysLysValValSerLeuLeuGlnValAsnGly 130
YY 1013 GGAACCTAAGGGGATTAATTAATGACACAGTAAGGTGGAAGGTTTCTTTCTTAAGCGAG 954
DB 131 -----TyrSerGly----- 133
YY 953 ACCAGAGTTTATTTATTTACTCAATCAGCTCTCCCA-AAAAGTTACTTTAAATGCTGTTT 895
DB 134 -----IleLeuLeuLeuAlaCysIleSerValAspArgTyrLeuAlaIleValIHis 150
YY 894 TCTTCCCACTTTATGAAACACATTTGCTGCTGAGACGGGAGCTATGCTATGCTCAAAAT 835
DB 151 AlaThrArg-----ThrLeuThrGlnLysArgTyrLeuValLysPhe 164
YY 834 ATTGT-----CTGTTGACATGACATTAAGTCTTTTAAAAAAA 796
DB 165 IleCysLeuSerLetyrPglyLeuSerLeuLeuAlaLeuProValLeuLeuPheArg 184
YY 795 AAACAGTGCCTTCAGATTAAGAAATTAAGGGCTTAATTAAGATTAATTAATGT 736
DB 185 ArgThrValTyrSerSerAsnValSerProAla-----CysTyrGluAsp----- 199
YY 735 CCTTTTGTGAGACAGAGCTCTGCTTTTAAACAAACAAAGAAATGAGAGATTCTA 676
DB 200 -----MetGlyAsnAsnThrAlaAsnTyrArgMetLeu 210
YY 675 CTACTGCTGCTGCTCTCCACAGGGGATTCATAGAGAT----- 637
DB 211 LeuArgIleLeuProGlnSerPheGlyPheIleValProLeuLeuIleMetLeuPheCys 230
YY 636 -----CSTAAAGAAAGAAACAACTGCGAGGGAGAGACGACTTGATTT 592
DB 231 TyrGlyPheThrLeuArgThrLeuPheLysAlaHisMetGlyGlnLysHisArgAlaMet 250
YY 591 CAGGTAAGTAACCAAAAAATTTAGCTATCTCTTTCGACAGACATGCTCTTCTCTGTT 532
DB 251 ArgValIlePheAlaValIleLeuIlePheLeuLeuGlyStrp-LeuProTyrAsnLeuVal 270
YY 531 AACACTTTCA-----GTCATTCAAAAGGCTTTTGCTACA 498
DB 270 IleLeuLeuAlaAspThrLeuMetArgThrGlnValIleGlnGlnThrCysGlnArg 288

```

RESULT 10
ID AAR80758 standard; Protein: 360 AA.
XX AAR80758;
XX
DT 26-MAR-1996 (first entry)
XX

DE Interleukin 8 receptor B.

XX KW Interleukin; IL-8; inflammation; psoriasis; dermatitis;
KW rheumatoid arthritis; inflammatory bowel disease;
KW chronic lung inflammation; treatment; antibody;
KW affinity purification; detection.

OS Homo sapiens.

XX PN US5440021-A.

XX PD 08-AUG-1995.

XX PF 29-MAR-1991; 91US-0677211.

XX PR 25-FEB-1994; 94US-0202056.

XX PR 29-MAR-1991; 91US-0677211.

XX PA (CHUN/) CHUNTHARAPAI A.
XX PA (HEBE/) HEBERT C.
XX PA (KIMK/) KIM K J.
XX PA (LEEJ/) LEE J.

XX PI Chuntharapai A, Hebert C, Kim KJ, Lee J;
XX DR WPI: 1995-283151/37.
XX DR N-PSDB; AAQ99008.

XX PT New antibodies against interleukin 8 type B receptor - used to treat
XX PT or prevent inflammation, also for detecting receptor expression and
XX PT purification.

XX PS Disclosure: Columns 53-56; 62pp; English.

XX CC Antibodies directed against the interleukin-8 receptor B can be used
XX CC to treat or prevent inflammation e.g. psoriasis, dermatitis,
XX CC rheumatoid arthritis and particularly inflammatory bowel disease and
XX CC chronic lung inflammation. When immobilised, these antibodies may
XX CC be used to detect interleukin-8 receptor B expression in cells and
XX CC tissues and for affinity purification of interleukin-8 receptor B
XX CC from cells.

XX SQ Sequence 360 AA:

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
0.344	85.50	37.84%	22.39%	3.70%	360	58	40	87	75	10

US-09-830-244B-2 (1-1331) x AAR80758 (1-360)

```

YY 1187 AGTCGATTTTGGCCATCAAGCCCTCAATATCATGCTCTGCTACTGATG-----ATA 1134
DB 76 SerArgValGlyArgSerValThrAspValTyrLeuLeuGlnLeuAlaLeuAlaAspLeu 95
YY 1133 TTTAAACGATATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1074
DB 96 LeuPheAlaLeuThrLeuProIleTyrPalaIaSerLysValAsnGlyTyrPheGly 115
YY 1073 AACTGCGTATAGGCTCTTGGTATAAAGCAATATCCAGAAATTAAGATTAATTAATTT 1014
DB 116 ThrPheLeu-----CysLysValValSerLeuLeuGlnValAsnGly 130
YY 1013 GGAACCTAAGGGGATTAATTAATGACACAGTAAGGTGGAAGGTTTCTTTCTTAAGCGAG 954
DB 131 -----TyrSerGly----- 133
YY 953 ACCAGAGTTTATTTATTTACTCAATCAGCTCTCCCA-AAAAGTTACTTTAAATGCTGTTT 895
DB 134 -----IleLeuLeuLeuAlaCysIleSerValAspArgTyrLeuAlaIleValIHis 150

```

QY	894	TCCTCCACCTTATATGAACACCATTTGCTACTGTTGACAGGGGACATATCTGATCAGAAAT	835
Db	151	AlaThrArg-----ThrLeuThrGlnLysArgTyrLeuValLysPhe	166
QY	834	ATTGTG-----CTGTGACATGACGACTTAACTGTTAAAAAA	766
Db	165	IleCysLeuSerIleTyrPheLysSerLeuLeuAlaLeuProValIleLeuPheArg	184
QY	795	AAACAGCGCCTTCAGATTAAGAAATTAGGGGCTCTAATTAAGATGTTAATGATAATGT	736
Db	185	ArgThrValIleTyrSerSerAsnValSerProIa-----CysTyrGluAsp-----	199
QY	735	CCTTTTTTTTGACAGACAGCTCTCTGTCTTTAAACAAACAAAGAAATGAGAGAGTTCTA	676
Db	200	-----MetLysAsnSerThrIleAsnTyrArgMetLeu	210
QY	675	CTAACTGCTCTGTGCTCCACAGAGGGATTCGAATGAGAT-----	637
Db	211	LeuArgIleLeuProGlnSerPheCylPheIleValProLeuLeuIleMetLeuPheCys	230
QY	636	-----CCTAAGAGGAAACAAACACCTGCAGGGGAGGAGACAGCATTTGGATTT	592
Db	231	TyrGlyPheThrLeuArgThrLeuPheLysAlaHisMetCylGlnLysHisArgAlaMet	250
QY	591	CAGATTAAGTACCAAAAAATATTATAGCTATCTTTCTGCAGACATCCCTCTTCTGTGT	532
Db	251	ArgValIlePheAlaValIleValIlePheLeuLeuCysTyr-LeuProTyrAsnLeuVa	270
QY	531	AACCTTTCA-----GTCAATCAAAAGCCTATGGTAGA	498
Db	270	IleuLeuAlaAspThrLeuMetArgThrGlnValIleGlnGlnThrCysGluArg	288
RESULT 11			
AAG80121			
XX	ID	AAG80121 standard; Protein; 360 AA.	
XX	AC	AAG80121:	
XX	DT	17-JAN-2002 (first entry)	
XX	DE	Human CXCR2 protein.	
KW	Chemokine; tumour diagnosis; colorectal; prostatic; organ rejection;		
KW	Inflammation; autoimmune disease; metastasis; bronchial asthma; lupus;		
KW	chronic bowel inflammation; rheumatoid arthritis; cytostatic;		
KW	antirheumatic; antiasthmatic; immunosuppressive; dermatological;		
XX	Homo sapiens.		
XX	OS		
XX	PN	WO200172830-A2.	
XX	PD	04-OCT-2001.	
XX	PF	02-APR-2001; 2001WO-EP03708.	
XX	PR	31-MAR-2000; 2000DE-1016013.	
PA	(IFPP-) IFF PHARM GMBH.		
PA	(FORS/) FORSMANN U.		
PI	Forssmann W, Adermann K, Heitland A, Spodsberg N;		
DR	WPI: 2001-626256/72.		
PT	Diagnostic agent containing two or more receptor-specific ligands,		
PT	useful for detecting tumors, inflammation etc., also therapeutic use of		
XX	ligand inhibitors		
XX	PS	Disclosure: Page 12; 26pp; German.	
XX	CC	This invention describes a novel diagnostic agent (A) comprising at least	

CC two different ligands (I) for receptors (II) that are implicated in
CC disease. (A) are used for the diagnosis of tumors (especially colorectal
CC or prostatic), organ rejection, inflammation and autoimmune diseases.
CC Also inhibitors of (I) are used therapeutically against tumors (and their
CC metastases), inflammation (particularly bronchial asthma or chronic bowel
CC inflammation), or autoimmune diseases (rheumatoid arthritis or lupus),
CC where the (cardio)vascular, lymphatic, respiratory, nervous, digestive,
CC endocrine, motor or urogenital systems or skin are affected, and bone
CC marrow diseases. The products of the invention are chemokine derivatives
CC which have cytostatic, antiinflammatory, antisthmatic,
CC immunosuppressive, dermatological, antihemetic, antidiabetic,
CC Chemokines act on specific tumor and inflammatory cells through a
CC costimulation of chemokine receptors (CR), which control migration and
CC proliferation of these cells. AA080045-AA080128 represent human chemokine
CC fragments used to illustrate the method of the invention.

XX Sequence 360 AA;

SQ

Alignment Scores:

Pred. No.:	Length:
Score: 0.344	360
Percent Similarity: 85.50	Matches: 58
Best Local Similarity: 37.84%	Conservative: 40
Query Match: 22.39%	Mismatches: 87
	Indels: 75
	Gaps: 10

US-09-830-244B-2 (1-1331) x AAG80121 (1-360)

OY 1187 AGTCGATTATTTGCCATCAAGCCTCATTAATTCATGGTCCGTACTATG-----ATA 1133
| | | | | : : : | | : : : : : | | | | : : :
Db SerArgValGlyArgSerValThrSpValTYrLeuLeuAsnLeuAlaLeuLysPheuu 95
OY 1133 TTTAAACGTAATATGTTGGTATTGGTGTTGGTATGGTTCGCCTTAGCGTTGCTTCGG 1074
| | | : : : | | : : : | | | : : : | | | : : :
Db LeuPheAlaLeuThyrLeuProIleTrpAlaIleSerLysValAsnGlyTYrIlePheGly 115
OY 1073 AACGCGCATATGAAGGCTCTTTGCTTAAGAACAATATCCAGAAATAGATAATTAACATT 1014
: : : | | : : : | | : : : | | : : : | | : : : | | : : :
Db ThrPheLeu-----CysLysValValSerLeuLeuLYSGlVAlaSnPhe 130
OY 1013 GSAACTAAGGGGATATMAATGACACAGTAAGGTGAGGTTTTTTTTTTCTTAAGGAG 954
131 -----TySerGly----- 133
OY 953 ACCAAGATTATTTATTTACTCAATCACTGCCA-AAAAGTTACTTTAAATGCTGTGTT 895
: : : | | | | | : : : | | : : : | | : : : | | : : :
Db -----IleLeuLeuAlaIcylsIleSerValIsparGYrLeuAlaIleValHis 150
OY 894 TCCTCCCACTTTAAGAACCACATTTGCTACTGTGGACAGGGCACTTCGTANTCAGAAAT 835
: : : : : : : : : | | | : : : | | | : : : : : : : :
Db AlaThrArg-----ThreThrgInLysArgTYrLeuValLysPhe 164
OY 834 ATTGT-----CTGTTCGACATCAGACTTAATCTGTTTAAAAAAA 796
| | | | | | | | | : : : : : : : : : | | : : :
Db IleCysLeuSerIleTrpGlyLeuSerLeuLeuAlaLeuProValIleuLeuPhearg 184
OY 795 AAAACAGTCCTTACAGATTAAGAATTAAGGGCTCTTAATAAGATGTTTAAATGATAATGT 736
: : : | | | | | : : : | | : : : | | | : : : | | : : :
Db ArgThryValTYrSerSerAsnValSerProAla-----CystyrGluAsp----- 199
OY 735 CCTTTTGGAGACAGAGTCTGTGCTTAAAAAACAAAAACAAGAAATGAGAGAGTTCTTA 676
200 -----MetGlyAsnAsnThrAlaSnTrpArgMetLeu 210
OY 675 CTAACTGTCCTGTGCTCCAGAGGGATTTCAATGAGAGAT----- 637
| | | : : : | | | | | : : : | | : : : | | : : : | | : : :
Db LeuArgIleLeuPheGlnSerPheGlyPheIleValIleProLeuLeuIleMetLeuPheCys 230
OY 636 -----CGTAAAGAAGAAACAACCTGACAGGGAAGCAGCAGCTTTGGATT 592
231 TyrcIyPheThrLeuArgIthrLeuPheLysAlaHisMetGlyGlnLysHisArgAlaMet 530
OY 591 CAGATAAAGTACCAAAATAATATTAGTATCTCTTTCTGACAGCATGCCCTTCTCTCTGTT 532

YY		894	TCTTCCACCATTTATGAACCACATTGTCACACTGTTCAGACGGGCATATCGATCAGAAAT	835
Dd		145	AAtHtArg-----ThLeuThrGlNlysArGtyrLeuValLysPhe	158
OY		834	ATTTGT-----CTGTGACATTCAGTAAGCTTTAAAAAAA	796
Dd		159	IleCyLSerSerIeTrpClyLeuSerLeuLeuAlaIleuProValLeuLeuPheArg	178
OY		795	AAACAGTGCCCTTCAGATAGAAATTAGGGCCTCAATTAAGATGTTATATGATAAATGT	736
Dd		179	ArgThValTyrrSerSerAsnValSerProAla-----CystyrGlunsp-----	193
OY		735	CCTTTTTTTTGAGACAGAGTCCTCTGTAAAAACAACAAGAAGAAATGGAGACTTCA	676
Dd		194	-----MetGlyAsnAsnThrAlaAsnTrpArgMetLeu	204
OY		675	CTAATGTCCTGTGCTCCACAGAGGAGATTCAATGAGGAT-----	637
Dd		205	LeuArgIleLeuProGlnSerPheGlyPheIleValProLeuLeuIleMetLeuPhecys	224
OY		636	-----CGTAAGAAGCAAAACAACCTGCAGGGGAGAACACACATCTTGATTT	592
Dd		225	TyrclyPheThrLeuArgThrLeuPheLysAlaHisMetGlyGlnLysHisArgAlaMet	244
OY		591	CAGATAAAGTACCAAAAATATTAGCTATCTCTTCGACAGATGCCCTCTTCTCTGT	532
Dd		245	ArgValIlePheAlaValIleValIlePheLeuLeucysTrp-LeuProTyraSnLeuVa	264
OY		531	AACATTTTCA-----GTCATTCAAAAGCCTATTGGTAGA	498
Dd		264	IleLeuAlaAspThrLeuMetArgThrGlnValIleGlnGlnThrCysGluArg	282
RESULT 14				
ID	AAB24943		standard; Peptide: 535 AA.	
XX	AAB24943:			
DT	27-NOV-2000	(first entry)		
DE	Plant SDF encoded polypeptide sequence SFO List 1 NO:474.			
KW	Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment;			
KM	SDF; genetic mapping; identification; promoter; structural gene; UTR;			
XX	untranslated region; expression control.			
OS	Plant.			
XX	MO200040695-A2.			
PN	13-JUL-2000.			
PD	07-JAN-2000; 2000MO-US00466.			
PF	08-JAN-1999; 98US-0115293.			
PR	(CERE-) CERES INC.			
PA	Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;			
PI	Zheng L;			
PP	WPI: 2000-465970/40.			
PRT	New corn plant and Arabidopsis thaliana sequence-determined DNA			
XX	fragments, useful for expressing gene products and for controlling			
XX	expression of a target gene -			
XX	Claim 14; Page 571-572; 673pp; English.			

CC as sequence-determined DNA fragments (SDPs), from corn plants and
 CC Arabidopsis thaliana. The SDPs are promoters, structural genes,
 CC untranslated regions (UTRs), or 3' termination sequences. They can be
 CC used for expressing a gene product and controlling expression of a
 CC target gene, either as a promoter, a structural gene, an UTR or as a
 CC 3' termination sequence. They are also useful as tools for genetic
 CC mapping, and identification of a particular individual plant or for
 CC clustering a group of plants with a common trait. AA78433 to AA78630
 CC and AAB24605 to AAB25099 represent the specifically claimed
 CC polynucleotide sequences and polypeptides encoded by them given in the
 CC present invention.

XX Sequence 535 AA:

Alignment Scores:
 Pred. No.: 0.468 Length: 535
 Score: 85.00 Matches: 38
 Percent Similarity: 44.06% Conservative: 25
 Best Local Similarity: 26.57% Mismatches: 56
 Query Match: 3.68% Indels: 25
 DB: 21 Gaps: 6

US-09-830-244B-2 (1-1331) x AAB24943 (1-535)

OY 1166 CTCATATATCATGCTCTG---CAGTTACTGATATTAAGCTATATTGTTGTA--- 1113
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 134 LeuSerSerHisIleLeuAlaAsnThrLeuValPheLeuProTyrIleuLeuIleAla 153
 OY 1112 -----TTGCGTGTGATAGTGTGTTGCGTATATGCTGTTGTTCTGG 1074
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 154 IleIleTyrSerValSerLeuTyrPheLeuValGlyLeuLeuSerPheSerTrpGlnAlaLeu 173
 OY 1073 AACTGCGTATAGGCTCTTGTGATTAAGCAATATCCAGAAATTAAGATATTAATTA 173
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 174 AlaTyrPheValLeuValIleTyrPheIleValLeuMetAlaAsnSerPheValLeuPhe 193
 OY 1013 GGAAGCTAA-----GGGATATTAATGACACAGTAAAGGTGGA 976
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 194 LeuSerSerLeuAlaProAsnTyrIleAlaGlyThrSerSerValThrIleLeuAla 213
 OY 975 GCGTTTTTTTTCTTAAGGAGACAGCTTTTATTATTAATCAATGAGCTCCCAAA 916
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 214 AlaPhePheLeuPheSerGly-----TyrPheIleSerLysGlnSerLeuProLys 230
 OY 915 AGTACTTAAATAGCTGTTCTTCCACTTATGAACACACATTTGCTACTGTTGACA 856
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 231 TyrTrpLeuPheMetTyrPhePheSerMetTyrLysTyrAlaLeuAspAlaLeuLeuIle 250
 OY 855 GGGAGCTAT---CTGATCAGAAATATTGCTGTTG-----ACATCAGTACTT 811
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 251 AsnGluTyrSerCysLeuHisAsnLysCysLeuValTrpPheGluGlnAlaSerValAsn 270
 OY 810 AACTGTTTA 802
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 271 SerCysLeu 273

RESULT 15

AAB24942 standard; Peptide: 575 AA.

AC AAB24942:

XX 27-NOV-2000 (first entry)

DE Plant SDF encoded polypeptide sequence SEQ List 1 NO:473.

KW Plant: corn; Arabidopsis thaliana; sequence-determined DNA fragment;

KM SDF: genetic mapping, identification; promoter; structural gene; UTR;

XX untranslated region; expression control.

OS Plant.

XX WO200040695-A2.

XX 13-JUL-2000.
 PD 07-JAN-2000; 2000WO-US00466.
 XX 08-JAN-1999; 99US-0115293.
 PF (CERE-) CERES INC.
 PR Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME.
 XX Zheng L;
 PI WPI; 2000-465970/40.
 DR
 XX
 XX
 PT New corn plant and Arabidopsis thaliana sequence-determined DNA
 PT fragments, useful for expressing gene products and for controlling
 PT expression of a target gene
 PS Claim 14; Page 569-570; 673pp; English.

The present invention describes polynucleotides, such as complete cDNA sequences and/or sequences of genomic DNA encompassing complete genes, portions of genes, and/or intergenic regions, collectively referred to as sequence-determined DNA fragments (SDPs), from corn plants and Arabidopsis thaliana. The SDPs are promoters, structural genes, and untranslated regions (UTRs), or 3' termination sequences. They can be used for expressing a gene product and controlling expression of a target gene, either as a promoter, a structural gene, an UTR or as a 3' termination sequence. They are also useful as tools for genetic mapping, and identification of a particular individual plant or for clustering a group of plants with a common trait. AA78433 to AA78630 and AAB24605 to AAB25099 represent the specifically claimed polynucleotide sequences and polypeptides encoded by them given in the present invention.

XX Sequence 575 AA:

Alignment Scores:
 Pred. No.: 0.482 Length: 575
 Score: 85.00 Matches: 38
 Percent Similarity: 44.06% Conservative: 25
 Best Local Similarity: 26.57% Mismatches: 56
 Query Match: 3.68% Indels: 25
 DB: 21 Gaps: 6

US-09-830-244B-2 (1-1331) x AAB24942 (1-575)

OY 1166 CTCATATATCATGCTCTG---CAGTTACTGATATTAAGCTATATTGTTGTA--- 1113
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 174 LeuSerSerHisIleLeuAlaAsnThrLeuValPheLeuProTyrIleuLeuIleAla 193
 OY 1112 -----TTGCGTGTGATAGTGTGTTGCGTATATGCTGTTGTTCTGG 1074
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 194 IleIleTyrSerValSerLeuTyrPheLeuValGlyLeuLeuSerPheSerTrpGlnAlaLeu 213
 OY 1073 AACTGCGTATAGGCTCTTGTGATTAAGCAATATCCAGAAATTAAGATATTAATTA 1014
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 214 AlaTyrPheValLeuValIleTyrPheIleValLeuMetAlaAsnSerPheValLeuPhe 233
 OY 1013 GGAAGCTAA-----GGGATATTAATGACACAGTAAAGGTGGA 976
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 234 LeuSerSerLeuAlaProAsnTyrIleAlaGlyThrSerSerValThrIleLeuAla 253
 OY 975 GCGTTTTTTTTCTTAAGGAGACAGCTTTTATTATTAATCAATGAGCTCCCAAA 916
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 254 AlaPhePheLeuPheSerGly-----TyrPheIleSerLysGlnSerLeuProLys 270
 OY 915 ACTTACTTAAATAGCTGTTCTTCCACTTATGAACACACATTTGCTACTGTTGACA 856
 DB |||:|||||:||||| |||: ||| :|||:|||||:
 DB 271 TyrTrpLeuPheMetTyrPhePheSerMetTyrLysTyrAlaLeuAspAlaLeuLeuIle 290
 OY 855 GGGAGCTAT---CTGATCAGAAATATTGCTGTTG-----ACATCAGTACTT 811
 DB |||:|||||:||||| |||: ||| :|||:|||||:

Mon Jun 30 08:44:15 2003

us-09-830-244b-2.n2p.rag

Page 13

Db	291	AsnGluTyrSerCysLeuHisAsnLysCysLeuValITrPheGluGluAlaSerValAsn	310
Oy	810	AACTGTTA	802
		:::	
Db	311	SerCysLeu	313

Search completed: June 24, 2003, 19:19:50
Job time : 72 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 24, 2003, 19:17:37 ; Search time 18.5 Seconds

(without alignments)

4233.720 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 2322

Sequence: 1 ctatgttttcaagaatcaag.....gccttcccaatgactctg 1331

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus.n2p.model -DEV=xlp
-Q=/cg22.1/USPnc_spool/US09830244/runat_24062003_130002_10088/app_query.fasta.1.1479
-DB=Issued_Patents_AA -QMT=fastn -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LISTEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=plco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -ALIGN=15
-USER=US09830244 -ECGN 1.1.28 -urnat_24062003_130002_10088 -NCPU=6 -ICPU=3
-NO_MMAP -LARGOQUERY -NEG_SCORES=0 -WAIT -DSPLLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Issued_Patents_AA:
1: /cg22_6/plodata/1/1aa/5A.COMB.pep.*
2: /cg22_6/plodata/1/1aa/5B.COMB.pep.*
3: /cg22_6/plodata/1/1aa/6A.COMB.pep.*
4: /cg22_6/plodata/1/1aa/6B.COMB.pep.*
5: /cg22_6/plodata/1/1aa/PCBUS.COMB.pep.*
6: /cg22_6/plodata/1/1aa/backfile1.pep.*

Pred: NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	93	4.0	384	4	US-09-071-035-276	Sequence 276, App
2	93	4.0	430	4	US-09-071-035-274	Sequence 274, App
3	85.5	3.7	360	1	US-08-202-056-7	Sequence 7, Appl
4	85	3.7	2391	2	US-08-446-855A-2	Sequence 2, Appl
5	85	3.7	2391	4	US-09-150-741-2	Sequence 2, Appl
6	84.5	3.7	355	1	US-07-759-568-1	Sequence 1, Appl
7	84.5	3.7	355	1	US-08-450-393A-8	Sequence 8, Appl
8	84.5	3.7	355	2	US-08-390-000A-5	Sequence 8, Appl
9	84.5	3.7	355	4	US-08-446-669-8	Sequence 8, Appl
10	84.5	3.7	355	5	PCR-US95-00476-8	Sequence 8, Appl
11	84	3.6	378	4	US-09-082-088-2	Sequence 2, Appl
12	84	3.6	378	4	US-09-546-117-2	Sequence 2, Appl

13	83.5	3.6	668	4	US-09-173-151A-35	Sequence 35, Appl
14	82.5	3.6	375	2	US-08-495-695B-28	Sequence 28, Appl
15	82.5	3.6	375	5	PCR-US94-14436-28	Sequence 18, Appl
16	79.5	3.4	2787	4	US-09-245-041-15	Sequence 15, Appl
17	77	3.3	615	4	US-09-134-001C-3284	Sequence 3284, Ap
18	76.5	3.3	350	1	US-08-202-056-1	Sequence 1, Appl
19	76.5	3.3	350	1	US-08-076-093A-2	Sequence 7, Appl
20	76.5	3.3	350	1	US-08-450-393A-7	Sequence 1, Appl
21	76.5	3.3	350	1	US-08-410-453A-1	Sequence 1, Appl
22	76.5	3.3	350	1	US-08-701-265-2	Sequence 2, Appl
23	76.5	3.3	350	1	US-08-410-454A-1	Sequence 1, Appl
24	76.5	3.3	350	2	US-08-284-586-2	Sequence 2, Appl
25	76.5	3.3	350	2	US-08-805-478-2	Sequence 2, Appl
26	76.5	3.3	350	2	US-08-802-627A-2	Sequence 2, Appl
27	76.5	3.3	350	2	US-08-801-238-2	Sequence 2, Appl
28	76.5	3.3	350	2	US-08-801-238-2	Sequence 2, Appl
29	76.5	3.3	350	3	US-09-104-296-2	Sequence 7, Appl
30	76.5	3.3	350	3	PCR-US95-00476-7	Sequence 7, Appl
31	76.5	3.3	350	5	US-09-029-267-2	Sequence 2, Appl
32	74.5	3.2	401	3	US-08-495-695B-32	Sequence 32, Appl
33	73.5	3.2	451	4	US-08-495-484-12	Sequence 12, Appl
34	73.5	3.2	2165	5	PCR-US95-12507-2	Sequence 2, Appl
35	73.5	3.1	485	2	US-08-477-451-22	Sequence 22, Appl
36	73	3.1	490	4	US-09-126-420A-16	Sequence 16, Appl
37	73	3.1	366	4	US-09-134-001C-4412	Sequence 4412, Ap
38	72.5	3.1	156	4	US-09-134-001C-4450	Sequence 4, Appl
39	72.5	3.1	432	2	US-08-644-034A-2	Sequence 12, Appl
40	72.5	3.1	1226	2	US-08-540-804-12	Sequence 12, Appl
41	72	3.1	1226	2	US-08-218-265-12	Sequence 12, Appl
42	71.5	3.1				
43	71.5	3.1				
44	71.5	3.1				
45	71.5	3.1				

ALIGNMENTS

RESULT 1
US-09-071-035-276
Sequence 276, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36, 373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 276:
SEQUENCE CHARACTERISTICS:

/	TELEX:	910/371-7168
/	INFORMATION FOR SEQ ID NO:	7:
/	SEQUENCE CHARACTERISTICS:	
/	LENGTH:	360 amino acids
/	TYPE:	amino acid
/	TOPOLOGY:	linear
/	US-08-202-056-7	

Alignment Scores:			
Pred. No.:	0.0534	Length:	360
Score:	85.50	Matches:	58
Percent Similarity:	37.84%	Conservative:	40
Best Local Similarity:	22.39%	Mismatches:	87
Query Match:	3.70%	Indels:	75
DB:	1	Gaps:	10

US-09-830-244B-2 (1-1331) x US-08-202-056-7 (1-360)			
QY	1187	AGTCGATATTGGCAATCAACGCCAATAATCATATGCTCGTACTGATG-----ATA	1133
Dd	76	SerArgValGlyIargSerValThrAspValTYrLeuLeuAlaSerLeuAlaAsnAspLeu	95
QY	1133	TTPAAACGTAATATGTCTGTTGCTATTCCTTGCTGGTAGTGCTTGCGTATAGTGTTGTTCTGG	1074
Dd	96	LeuPheAlaLeuThrLeuProIleTrpAlaAlaSerLysValAsnGlyTrpIlePheGly	115
QY	1073	AACCTGCATGAAGAAGCCTCTGTGGTATTAAGAACAATACAGAAATTAAGTAACTTT	1014
Dd	116	ThrPheLeu-----CysLysValValSerLeuLeuLysGluValAsnSph	130
QY	1013	GGAACTAAGGCGATTAATAATGACACAGTAAGAAGTGAAGGCTTTTTTTCTTAGGGAG	954
Dd	131	-----TyserGly-----	133
QY	953	ACCAGAGTTTATTAATTAATCAATCACATCCCTCCA-AAAGTTACTTTAAATGCTCTTT	895
Dd	134	-----IleuLeuLeuAlaIacysIleSerValAspArgTYrLeuAlaIleValHis	150
QY	894	TCCTCCCACCTTTATGTAACAACACATTGGCTACTGGTGGACAGGAGCATCTTCATGACAAAT	835
Dd	151	AlaThrArg-----ThirLeuThrGlnLysArgTYrLeuValLysPhe	164
QY	834	ATTGTG-----CTGTGCACATCAGTACTTAACCTGTTAAAAAAA	796
Dd	165	IleCysLeuSerIleTrpGlyLeuSerLeuLeuAlaLeuProValLeuLeuPheArg	184
QY	795	AAAACAGTGCCTCAGATTAAGAATAATGAGGCGCTCAATTAAGATGTTTAATGATAAATGT	736
Dd	185	ArgThrValTYrSerSerAsnValSerProAla-----CysTYrGluAsp-----	199
QY	735	CCTTTTTTTTGAACAGAGCTCTGTCTTTAAAAAACAAACAAAGAAATGAGAGGTCTTA	676
Dd	200	-----MetGlyAsnAsnThrAlaAsnTrpAlaSerTrpAlaMetLeu	210
QY	675	CTAATGTCCTGTGCTCTCCACAGAGGAGATTCAATGAGAT-----	637
Dd	211	LeuArgIleLeuProGlnSerPheGlyPheIleValProLeuLeuIleMetLeuPheCys	230
QY	636	-----CGTAAGAAGAGAAACACCTCAGGAGGAGAGCAGCATCTTGATTT	592
Dd	231	TyrOlyPheThrLeuArgThrLeuPheLysAlaHisMetCylGlnLysHisArgAlaMet	250
QY	591	CAGTAAAGTAAACCAAAATATATAGCATGCTCTTTGSCAGACATGCGCTTCTCTGTT	532
Dd	251	ArgValIlePheAlaValAlaLeuIlePheLeuLeuCysTrp-LeuProTYrAsnLeuVa	270
QY	531	AACACTTTCA-----GTCATTTCAAAGGCGATTCGTAGA	498
Dd	270	IleuLeuAlaAspTrpLeuMetArgThrGlnValIleGlnGlnLysTrpCysGluArg	288

RESULT 4
 US-08-446-855A-2
 Sequence 2, Application US/08446855A

QY	619	CCTCAGGGGAGAGACATCTGGATTTCAGATTAAGTAC-----CAAAAATATTAGC	566
Db	677	-----GlnTyrArgIuIuIuLeuGlyIuYraSnIleIleTylIleLysLysValleuLe	695
QY	565	TATCTCTTTCGACAGACATGCCCTCTTCTCTGTTAAACACTTTCAGTCATTCAAAAGCTA	506
Db	695	uLeuGlySerGlyGlyLeuGlyIleGlyGlnIleGlyIuPhaAspTyrSer---Glyh	714
QY	505	TTGGTAGAGGTGGACGAATCCAAAGTGATATTACTGTACAGTTTATTAACTATGCT	446
Db	714	rgInAlaIleLysSerIleuLysGluLys-----GlyIleTyr-ValIleuVa	730
QY	445	GTATTCACCAATTAACGGCATTTACCAATTAATGCGATTAGTCATCAAAAAGATCGAAA	386
Db	730	IaSnProaSnIleAlaIleThrVal-Gln-----ThrSerLysGlyLeuAlaA	745
QY	385	TGCTATGGGACACTTTCACCAAAAAGAAAAGAAATCCAATGCGCTCTATAAAGGAGANT	326
Db	745	sPluYsAllyTrpHeIeuProValIaSnCysGluPhelValGluLysIleIleLysGluL	765
QY	325	ATAGCAAT-----GGCAGACGAGAAATGAATTTCTGT-----	294
Db	765	ysProAspHeIleLeuGlyStnPhneGlyGlyGlnThrAlaLeuAsnCysAlaLeuMetL	785
QY	293	-----CAAAAGAAAGTACTTCTCAAAAGCAATCAGAGAAAACGTGTCCAAATTAGCGCTG	239
Db	785	euaSpIuIuLysLysValleuLysLys-----AsnAsnCysGlnCysLeuGlyThr-	801
QY	238	GGAGGACCATTAAGCCACATATCATGATATTATTTGTTTGACCTTAGGGGCCCTCA	179
Db	802	-----SerLeuGlnSerIleA	807
QY	178	ATATTCCCTCATCTGCTCAAACTTCAAACTGTTTACTGCACAGATGTGATTAATTCAGC	119
Db	807	rgIleThrGlu-----AsnArgThrIleuPhe-----AlaGluLysLeuLysG	821
QY	118	CAAGTAACCTTAACATY---CCCTATGGCAGTGCAGAAA	84
Db	821	IuIleAsnGluArgIleAlaProTyrGlySerAlaLys	853
RESULT 5.			
US-09-150-741-2			
/ Sequence 2, Application US/09150741			
/ Patent No. 6183996			
/ GENERAL INFORMATION:			
/ APPLICANT: Stewart et al.			
/ TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate			
/ Patent No. 6183996			
/ TITLE OF INVENTION: Synthetase II			
/ FILE REFERENCE:			
/ CURRENT APPLICATION NUMBER: US/09/150,741			
/ CURRENT FILING DATE: 1998-09-10			
/ EARLIER APPLICATION NUMBER: PL6380			
/ EARLIER FILING DATE: 1997-12-16			
/ EARLIER APPLICATION NUMBER: A093/00617			
/ EARLIER FILING DATE: 1993-12-02			
/ EARLIER APPLICATION NUMBER: 08/446,855			
/ EARLIER FILING DATE: 1995-07-06			
/ NUMBER OF SEQ ID NOS: 15			
/ SOFTWARE: PatentIn Ver. 2.0			
/ SEQ ID NO 2			
/ LENGTH: 2391			
/ TYPE: PRT			
/ ORGANISM: Artificial Sequence			
/ FEATURE:			
/ OTHER INFORMATION: Description of Artificial Sequence: protein			
US-09-150-741-2			
Alignment Scores:			
Pred. No.:	0.159	Length:	2391
Score:	85.00	Matches:	82
Percent Similarity:	38.85%	Conservative:	40

Best Local Similarity: 26.118
Query Match: 3.688
DB: 4
Mismatches: 110
Indels: 83
Gaps: 18

US-09-830-244B-2 (1-1331) x US-09-150-741-2 (1-2391)

```

QY 930 ATCAGTCTCCCAAAAGTTACTTAAATGCTGTTTCTCCCACTTATGAAACCAT 871
DB 572 IIESErLeuGlyCyAspThrLySmetLySrgLYAsnAsgLYValAsnGlnPro 591
QY 870 TTGCTACGTGTGACAGGGGACATCTGATCAGAAATATTTCTGTTCACATCAGTACTT 811
DB 592 ValIle-----GlnLeuValAspAsnIleCysTyrIleThrSerGlnAsn 606
QY 810 AAC-----TGTTTAAAAAAAACAGTCCCTCAGATAGAAATTAAGGGCTCTATA 757
DB 607 HIsGlyTyrCysLeuLYsLYsLYsSerIleLeuLYsArgLYsGlnLeuAlaIleSerTyr 626
QY 756 AGATCTTAAATGATTAATGCTCTTTTGTGAGACAGAGTCTCTCTTAAACAA 697
DB 627 ILeAsnAlaAsnAspLYsSer-Ile-----GlnGlyIleSer----- 638
QY 696 ACAAGAAATGAGAGATCTACTACTCTCTCTGCTCCAGAGGGGATTCAT---GA 641
DB 639 -HisLYsAsnGlnLYsArgPheTyrSerValGlnPheHisProGlnGlyAsnAsnGlyProGln 658
QY 640 GGAT-----CGTAGAGAGAGAAACAA 620
DB 658 uAspThrSerPheLeuPheLYsAsnPheLeuLeuAspIlePheAsnLYsLYs----- 676
QY 619 CCTGCGAGGAGAGACATCTTGATTTTCAGATTAACATAC-----CAAAATATTAGC 566
DB 677 -----GlnTyrArgGlnTyrLeuGlnTyrAsnIleIleTyrIleLYsLYsValLeuLeu 695
QY 565 TATCTCTTTCTGAGACATCCTCTTCTCTGTTACACTTTCAGATTCATCAAAAGCTTA 506
DB 695 uLeuGlySerGlyLYsLYsLYsIleGlyGlnAlaGlyLupheAspTyrSer---GlyTh 714
QY 505 TTGGTAGAGGTGACGCAATTCAGATGTAGATTACTGTACAGATTATTAATCTGCT 446
DB 714 rGlnAlaIleLYsSerLeuLYsGlnCys-----GlyIleTyrValIleLeuVal 730
QY 445 GTATTCCACCAATTAAGCGCATTAACCATTAATGGCATTAAGTCACAAAGATCGMAA 386
DB 730 LAsnProAsnIleAlaThrVal-Gln-----ThrSerLYsGlyLeuAla 745
QY 385 TGCTAATGGACATTTCCCAAAAGAAAGAAATCCATGCTCCTATTAAGGAGAT 326
DB 745 sPLYsValTyrPheLeuProValAsnCysGlnPheAlaGlnLYsIleIleLYsLYsGln 765
QY 325 ATAGCAAT-----GGCAGCAGAAATGMAAATTTCTGT----- 294
DB 765 ySProAspPheIleLeuCYsThrPheGlyGlyGlnThrAlaLeuAsnCYsAlaLeuMetL 785
QY 293 -----CAAAAGAAAGTACTTCTCAAGCAATGAGAAACACTGTTCCCAATTAAGCCTG 239
DB 785 euAspGlnLYsLYsValLeuLYsLYs-----AsnAsnCysLYsLYsLeuGlyThr- 801
QY 238 GGGAGCCACATTAAGCCACTATCTCAGATAGTATGATTTGTTGTTGGCTTGGGGCCCA 179
DB 802 -----SerLeuGlnSerIleA 807
QY 178 ATATTCCCTCATATCTGCTAAACTCAACTGTTTACTCAGACAGTGTATTAATTCAGC 119
DB 807 rGlnThrGln-----AsnArgThrLeuPhe-----AlaGlnLYsLeuLYsG 821
QY 118 CAAGTAACTTAACAT---CCCTATGAGCAGTGAAGAA 84
DB 821 LuIleAsnGlnLYsArgIleAlaProTyrGlySerAlaLYs 833

```

```

? Patent NO. 5374506
? GENERAL INFORMATION:
? APPLICANT: Murphy, Philip M.
? TITLE OF INVENTION: Cloning of cDNA Encoding a Functional
? TITLE OF INVENTION: Human Interleukin-8 Receptor
? NUMBER OF SEQUENCES: 5
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Cushman, Darby & Cushman
? STREET: 1615 L Street, N.W.
? CITY: Washington
? STATE: D.C.
? COUNTRY: U.S.A.
? ZIP: 20036-5601
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/07/759,568
? FILING DATE: 19910913
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Scott, Watson T.
? REGISTRATION NUMBER: 26581
? REFERENCE/DOCKET NUMBER: MTS/5683/91535/MBH
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-861-3000
? TELEFAX: 202-822-0944
? TELEX: 6714627 cush
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 355 amino acids
? TYPE: AMINO ACID
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: peptide
? US-07-759-568-1

Alignment Scores:
Pred. No.: 0.0713 Length: 355
Score: 84.50 Matches: 58
Percent Similarity: 37.84% Conservative: 40
Best Local Similarity: 22.39% Mismatches: 87
Query Match: 3.65% Indels: 75
DB: 1 Gaps: 10

US-09-830-244B-2 (1-1331) x US-07-759-568-1 (1-355)
QY 1187 AGTCGATTTCGCAATCAAGCCCTCAATATCATGCTCTGCGACTTACTGATG---ATA 1134
DB 71 SerArgValGlnArgSerValThrAspValTyrLeuLeuAsnLeuAlaLeuAspLeu 90
QY 1133 TTTAAAGTATATTGTTGGTATTTGGTATTTGGTATTTGGTATTTGGTATTTGGTATTTGGT 1074
DB 91 LeuPheAlaLeuThrLeuPhePheAlaAlaSerLYsValAsnGlyTyrPheGly 110
QY 1073 AACTGCTATTAAGCGCTTGTGATTAAGCAATTCAGAAATTAAGTATTAATTAATT 1014
DB 111 ThrPheLeu-----CysLYsValValSerLeuLeuLYsValValAsnPhe 125
QY 1013 GGAACATAAGGGATTAATTAATGACACAGTAAGGTGAAGCGTTTCTTAAAGCGAG 954
DB 126 -----TyrSerGly----- 128
QY 953 ACCAGATTATTAATTAATCAATCAGTCCCA-AAAAGTTACTTAAATGCTGTTT 895
DB 129 -----IleLeuLeuLeuAlaCysIleSerValAspArgTyrLeuAlaIleValHis 145
QY 894 TCTTCCACTTATTAAGAACACATTTGCTGCTGAGCAGGGGACTATCTGATCAGAAAT 835
DB 146 AlaThrArg-----ThrLeuThrGlnLYsArgTyrLeuValLYsPhe 159

```

RESULT 6
US-07-759-568-1
Sequence 1, Application US/07759568

ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York

STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/390,000A
FILING DATE: 17-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6923-052
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-390-000A-5

Alignment Scores:
Pred. No.: 0.0713
Score: 84.50
Percent Similarity: 37.848
Best Local Similarity: 22.398
Query Match: 3.658
Length: 355
Matches: 58
Conservative: 40
Mismatch: 87
Indels: 75
Gaps: 10

US-09-830-244B-2 (1-1331) x US-08-390-000A-5 (1-355)

1187 AGTGTATTGCCAATCAAGCTCAATATCATCTCCGAGTACTGATG-----ATA 1134
71 SerArgValGlyArgSerValThrAspValTyrLeuLeuLysLeuAlaLeuAlaAspLeu 90
1133 TTAACGTAATATGTTGGTATGTTGGTATGTTGGTATGTTGGTATGTTGGTATGTTGG 1074
91 LeuPheAlaLeuThrLeuProIleThrPalaIleSerLysValAsnGlyTrpIlePheGly 110
1073 AACTGGCTAAGAGCTCTTGGTATTAAGCAATATCCAGAAATAGATAATTAATTAATTA 1014
111 ThrPheLeu-----CysLysValValSerLeuLeuLysGluValAsnPhe 125
1013 GGAACCTAAGGGGATTAATGACACAGTAAGGTGGAAGGTTTCTTCTTAAGGAG 954
126 -----TyrSerGly----- 128
953 ACCAGAGTTTATTATTACTCAATACAGTCTCCCA-AAAAGTTACTTTAAATGCTGTT 895
129 -----IleLeuLeuLeuAlaCysIleSerValAspArgTyrLeuAlaIleValIAs 145
894 TCTTCCCACTTATATGAACACATTGTCATCTGACAGGAGGAGCATGATCAGAAAT 835
146 AlaThrArg-----ThrLeuThrGlnLysArgTyrLeuValLysPhe 159
834 ATTTGT-----CTTTGACATCAGTACTTAAGTATTAAATAAA 796
160 IleCysLeuSerIleThrPylLeuSerLeuLeuAlaLeuProValLeuLeuPheArg 179
795 AAAACAGTCCCTTCAGATTAAGAAATTAAGGCTCTAATAAGTATTAAATGAATG 736
180 ArgThrValTyrSerSerAsnValSerProAla-----CysTyrGluAsp----- 194
735 CTTTTTTTTGAGACAGAGCTCTGCTTTAAAAAACAACAAGAAATGAGAGTTCTTA 676
195 -----MetGlyAsnAsnThrAlaAsnTrpArgMetLeu 205

675 CTAAGTNCCTGCTGCTCCACAGGAGGATTCATAGAT----- 637
206 LeuArgIleLeuProGlnSerPheGlyPheIleValProLeuLeuIleMetLeuPheCys 225
636 -----CCTAAGAACGAAACACACTGCAGGGGAAAGCAGCATCTTGATTT 592
226 TyrGlyPheThrLeuArgThrLeuPheLysAlaIleMetGlyGlnLysHisArgAlaMet 245
591 CAGATAAGTACCAAAATATTAAGTATCTCTTTCGACAGATGCGCTCTTCTGCT 532
246 ArgValIlePheAlaValValLeuIlePheLeuLeuCysTrp-LeuProTyrAsnLeuVal 265
531 AACACTTCA-----GTCATTCAAAAGCCTATTGCTAGA 498
265 IleuLeuAlaAspThrLeuMetArgThrGlnValIleGlnGluThrCysGluValArg 283

US-08-446-669-8
Sequence 8, Application US/08446669
Patent No. 6132987
GENERAL INFORMATION:
APPLICANT: Charo, Israel
TITLE OF INVENTION: MAMMALIAN MONOCYTE CHEMOATTRACTANT
TITLE OF INVENTION: PROTEIN RECEPTORS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306-2155
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,669
FILING DATE: May 25, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Nealey, Richard
REGISTRATION NUMBER: 30,092
REFERENCE/DOCKET NUMBER: UCAL-237/01US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5000
TELEFAX: 415-857-0663
TELEX: 380816COOLEYPA
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-446-669-8
Alignment Scores:
Pred. No.: 0.0713
Score: 84.50
Percent Similarity: 37.848
Best Local Similarity: 22.398
Query Match: 3.658
Length: 355
Matches: 58
Conservative: 40
Mismatch: 87
Indels: 75
Gaps: 10
US-09-830-244B-2 (1-1331) x US-08-446-669-8 (1-355)
1187 AGTGTATTGCCAATCAAGCTCAATATCATCTCCGAGTACTGATG-----ATA 1134
71 SerArgValGlyArgSerValThrAspValTyrLeuLeuLysLeuAlaLeuAlaAspLeu 90

OY	1133	TTAAAAACGTAATNTGTTGGTATTGGTTGGTGGATGGTGTGGCTTACGCTTACGCTGTTCTGG	10174
Db	91	LeuPheAlaLeuThrLeuProIleThrPalaIaLeuSerLysValaSnGlyThrIlePheGly	110
OY	1073	MACTGGCTGATTAAGGCTCTTTGGTTAAAGCAATATCAGCAAAATTAAGATTAATTAAGTT	10144
Db	111	ThrPheLeu-----CysLysValaLeuSerLeuLeuLysGluValaSnPhe	125
OY	1013	GGAAGTAAGGGATATTAATGACACAGTAAGAGTGGAAGGTTTTTTTCTTAAGGGAG	954
Db	126	-----TyrSerGly-----	128
OY	953	ACCAAGTATTATTAATTACTCAATACAGTCTCCCA-AAAGTACTTATTAAGTGTGTT	895
Db	129	-----IleLeuLeuLeuAlaCysLysSerValaIaPargTyrLeuAlaIleValHis	145
OY	894	TCTTCCCACTTTATATGAACACACATTGGTCTACTGTTGACAGGGAGCATATGCAGAAAT	835
Db	146	AlaThrValg-----ThrLeuThrGlnLysArgTyrIleValLysPhe	159
OY	834	ATTGT-----CTGTTCAGCATCAGTACTTAATCTGTATTAAGAAA	796
Db	160	IleCysLeuSerIleTrrPglyLeuSerIleuLeuAlaLeuProValLeuLeuPheArg	179
OY	795	AAACAGCGCTTCAGATTAAGAAATTAAGGCGCTTAATAGATGTTAATAGATAAATGT	736
Db	180	ArgThrValTyrSerSerAsnValSerProIa-----CysTyrGluAsp-----	194
OY	735	CCTTTTTTTTGAGACAGAGCTCTCTGTATAAAACAAACAAAGAAATGAGAGCTCTA	676
Db	195	-----MetCysIaSnAsnThrAlaIaSnTrrPargMetLeu	205
OY	675	CTAAGTGCCTGTGCTCCACAGAGGGATTCAGTAGAT-----	637
Db	206	LeuArgIleLeuProGlnSerPheGlyPheIleValProLeuIleIleMetLeuPheCys	225
OY	636	-----CGTAAGCAAGAAACAACTGCAGGGGAGAGCAATCTTGGAATT	592
Db	226	TyrGlyPheThrLeuArgThrLeuPheLysAlaHisMetCysGlnLysHisIaPargAlaMet	245
OY	591	CAGATTAAGTACCAAAAATTTAGCTATCTTTTCGACAGACATGCCCTTTCTCTGT	532
Db	246	ArgValIlePheAlaValaIleuIlePheLeuLeuCysTrr-LeuProTyrAsnLeuVa	265
OY	531	AACCTTTCA-----GTCATCAAAAGCTTATGGTGA	498
Db	265	IleLeuAlaIaAspThrLeuMetArgTrrGlnValIleGlnGlnThrCysGlnIuArg	283

RESULT 10
PCT-US95-00476-8
Sequence 8, Application PC/TUS9500476
GENERAL INFORMATION:
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: MAMMALIAN MONOCYTE CHEMOTACTIC ANTIGEN
TITLE OF INVENTION: PROTEIN RECEPTORS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Robbins, Berliner & Carson
STREET: 201 N. Figueroa Street, 5th Floor
City: Los Angeles
STATE: California
COUNTRY: USA
Zip: 90012-2628
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentia Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00476
FILING DATE:
CLASSIFICATION:

```

1  ATTORNEY/AGENT INFORMATION:
2  NAME: Berliner, Robert
3  REGISTRATION NUMBER: 20,121
4  REFERENCE/DOCKET NUMBER: 5555-2911
5  TELECOMMUNICATION INFORMATION:
6  TELEPHONE: 310-977-1001
7  TELEFAX: 310-977-1003
8  TELEX:
9  INFORMATION FOR SEQ ID NO: 8:
10 SEQUENCE CHARACTERISTICS:
11 LENGTH: 355 amino acids
12 TYPE: amino acid
13 TOPOLOGY: linear
14 MOLECULE TYPE: protein
15 HYPOTHEetical: NO
16 PCT-US95-00476-8

```

Alignment Scores:

Pred. No.:	0.0713	Length:	35
Score:	84.50	Matches:	58
Percent Similarity:	37.84%	Conservative:	40
Best Local Similarity:	22.39%	Mismatches:	87
Query Match:	3.65%	Indels:	75
DB:	5	Gaps:	10

US-09-830-244B-2 (1-1331) X PCT-US95-00476-8 (1-355)

OY	1187	AGTCGATATTTGGCAATCAAGCGCTCAATATATATGCTCCGAGTACATGCAAT-----ATA	1134
	71	SeRatgVAlGlyARgSerValThraspAllyrLeuLeuSnbLeuAlaLeuAlaAspLeu	90
OY	1133	TTAAACGATATATTTGCTGGTATTTGGTGGTATAGCTTTGCCGATATAGTGGTTCGCG	1074
	91	LeupheAlaLeuThrLeuProileThrAlaAlaSerLysValaAsnGlyTrilepHeGly	110
OY	1073	AACGCGCATAAAGGCTCTTGGTATATAAACCATATCCACAAATTAAGATATTAACCTT	1014
	111	ThrHeLeu-----CysLysValValSerLeuLeuLysGluValaAspHe	125
OY	1013	GGACSTAAGGGGATATTAATATACACACAGTAAAGGTGGAAGGCTTTTTTTCTTAAGGAGC	954
	126	-----TyrSerGly-----	128
OY	953	ACCAGATTTTATTTACTCATCAATCAGTCCCA-AAAAGTTACTTTAAATGCTGTT	895
	129	-----IleLeuLeuLeuAlaCysIleSerValAspArgLysLeuAlaIleValHis	145
OY	894	TCTTCCCACTTTATGAACCAACATTTCTCTACTGTTCGACAGGGCAGTATGATCAGAAAT	835
	146	AlaThrArg-----ThrLeuThrGlnLysArgTyrLeuValIlysPhe	159
OY	834	ATTGT-----CTGTTCGATCAGTACTTAACTGTTTAAAAAAA	796
	160	IleCysLeuSerIleTyrGlyLeuSerLeuLeuLeuAlaLeuProValIleLeuPheArg	179
OY	795	AAACGATGCTTCAGATGAAGAATTAGGGGCTCTAATAAGATGTTAATGAATTAATGT	736
	180	ArgThrValTyrSerSerAsnValSerProAla-----CysTyrGluAsp-----	194
OY	735	CGTTTCTTTTGAGACAGATCTCGTCTTAAAAACAAACAAAGAAATGAGAGTCTTA	676
	195	-----MetGlyAsnAsnThrIleAsnTrpArgHeLeu	205
OY	675	CTAATCTCTCTGTCCTCCAGAGGGGATTCATAGGAGAT-----	637
	206	LeuArgIleLeuProGlnSerPheGlyPheIleValProLeuLeuIleMetLeuPheCys	225
OY	636	-----TyrGlyPheThrLeuArgThrLeuPheLysAlaHisMetCysIleLysHisArgAlaMet	592
	226	CAAGTAAAGTACCAAAATATATTAGTATCTTTCTGACAGACATGCGCTTTTCTGTT	532

[illegible]

```

QY 836 TTTCGATCAGATAGTCCCTGTCACAGTACGAAA----- 871
Db 139 ThMeIleuysrProtyrAspAlaenLysArgHisArgValPheLeuLeuIle 158
QY 872 -----TGTCGTTTCATTAAGTGGAGAGAAACAGCATTTTAACTATTCTTTGGGGA 925
Db 159 GlyMeCysTrpLeuIleAlaPheThrLeuGlyAlaLeu--ProIleuGlyTrpSnc 178
QY 926 CT-----GATTGAGTATTAATAAACTGTCGTCCTTGAAGAAAAAACCTTCC 979
Db 178 ySLeuhIsnLeuProAspCysSerThrIleLeuProLeuTyrSerLysLys--TyrI 197
QY 980 ACCTTAGTGTCTC---ATTATATCCCTAGTCCAACTATATTTCTTATTTCTGC 1036
Db 197 leAlaPheCysIleSerIlePheThrAlaIleLeuValThrIleValIleLeuTyrAla 217
QY 1037 ATATTGCTTTTATACCAAGAGCCTTATACAGCTTCAGACACCAACCATATACGACA 1096
Db 217 rGileTyPheLeuValLysSerSerSerArgLysValAlaAsn-----HisA 233
QY 1097 ACCATACCAACCAATACCAACAATACCTTTTATATATC 1140
Db 233 snAsnSerGluArgSerMetAlaLeuLeuArgThrValIle 247

```

```

RESULT 13
US-09-173-151A-35
Sequence 35, Application US/09173151A
Patent No. 6326472
GENERAL INFORMATION:
APPLICANT: Timans, Jacqueline C.
APPLICANT: Debels, Johannes Eduard Maria
APPLICANT: Antonius
APPLICANT: Sana, Theodore R.
APPLICANT: Bazan, J. Fernando
APPLICANT: Kastelein, Robert A.
TITLE OF INVENTION: Human Receptor Proteins; Related Reagents and Methods
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESS: DNA Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/173,151A
FILING DATE: 14-OCT-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/065,776
FILING DATE: 17-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/078,008
FILING DATE: 12-MAR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/081,883
FILING DATE: 15-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/095,987
FILING DATE: 10-AUG-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/078,416
FILING DATE: 18-MAR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/062,066
FILING DATE: 15-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.

```

```

REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0767X
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)852-9196
TELEFAX: (650)496-1200
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 668 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-173-151A-35
Alignment Scores:
Pred. No.: 0.131
Score: 83.50
Percent Similarity: 33.83%
Best Local Similarity: 22.89%
Query Match: 3.61%
DB: 4
Gaps: 11
Length: 668
Matches: 46
Conservative: 22
Mismatch: 50
Indels: 83

```

```

US-09-830-244B-2 (1-1331) x US-09-173-151A-35 (1-668)
QY 1155 TGTCTCGAGTACTGATGATTAATAACGTATATTTGGTATGCTGATGCTT 1096
Db 3 CysThrAspTrpSerIleAspIleLysLysTyr-----GlnValLeuValGlyLupPro 20
QY 1095 GTGCTATAGTGTGCTGCTG-----GAACTGCTGAT 1063
Db 21 ValArgIleLysCysAlaLeuPheThrIleArgThrAsnTyrSerLeuAlaGln 40
QY 1062 AAGGCT-----CTTGGTATTAAGCAATATCCAGAAATAGATTAATTAAT 1015
Db 41 SerAlaGlyLeuSerLeuMetTrpTyrLysSer----- 52
QY 1014 TGGAACTAAGGGCATATAATATGACACAGTAAGGTGGAAGGCTTTTCTTAAAGGA 955
Db 53 ---GlyProGlyAspPheGluGluProIleAlaPheAspGlySerArgMetSerLysGlu 71
QY 954 GACGAGATTTTATTATTACTCAATCAGTCTCCCAAAAGTTACTTAAATGCTGTT 895
Db 72 GluAspSerIle----- 75
QY 894 TCTTCCACTTATTAAGAACACATTTGCTACTGTGACAGGAGCTATCTGATCAGAAAT 835
Db 76 -----TrpPheArgProThrLeuLeuGlnAspSerLysLeuTyr----- 88
QY 834 ATTGTCTGTTGACATCAGTACTTAACTGTTTAAA-----AAAAACAGTG--- 787
Db 89 AlaCysValIleArgAsnSerThrTyrCysMetLysValSerIleSerLeuThrValGly 108
QY 786 CTTTCAGATAGAATAATAGGGCTTAATAGATTAATAGATTAATAGATTAATAGAT 727
Db 109 GluAsnAspThrGlyLeu-----CysTyrAsnSerLysMetLysTyrPhe 123
QY 726 TGAGACAGAGTCTCTGCTTAAATAAACA----- 697
Db 124 ---GluLysAlaGluLeuSerLysSerLysGluIleSerCysArgAspIleGluAspPhe 142
QY 696 -----ACAAAGAAATG 685
Db 143 LeuLeuProThrArgGluProGluIleLeuTyrTrpLysGluCysArgThrLysThrTrp 162
QY 684 AGA 682
Db 163 Arg 163

```

```

RESULT 14
US-08-495-695B-28
Sequence 28, Application US/08495695B
Patent No. 5976814
GENERAL INFORMATION:
APPLICANT: Bard, Jonathan A.

```

```

1  APPLICANT: Walker, Mary
2  APPLICANT: Brancheek, Theresa
3  APPLICANT: Weinschank, Richard L.
4  TITLE OF INVENTION: DNA ENCODING A HUMAN NEUROPEPT
5  TITLE OF INVENTION: Y/PEPTIDE YV/PANCREATIC POLYPE
6  TITLE OF INVENTION: RECEPTOR (Y4) AND USES THEREOF
7  NUMBER OF SEQUENCES: 33
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE: Cooper & Dunham
10 STREET: 1185 Avenue of the Americas
11 CITY: New York
12 STATE: New York
13 COUNTRY: U.S.A.
14 ZIP: 10036
15
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: PatentIn Release #1.0, Version #1.24
21
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/495,695B
24 FILING DATE: 13-Jan-1997
25 CLASSIFICATION: 435
26 ATTORNEY/AGENT INFORMATION:
27 NAME: White, John P.
28 REGISTRATION NUMBER: 28,678
29 REFERENCE/DOCKET NUMBER: 44743-A-PCT-US/JPW/JHB
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: (212) 278-0400
32 TELEFAX: (212) 391-0525
33 INFORMATION FOR SEQ ID NO: 28:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 375 amino acids
36 TYPE: amino acid
37 TOPOLOGY: linear
38
39 MOLECULE TYPE: protein
40
41 US-08-495-695B-28

```

REGISTRATION NUMBER: 28, 678
 REFERENCE/DOCKET NUMBER: 44743-A-PCT\JP\MAT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0525
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 375 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US94-14436-28

Alignment Scores:
 Pred. No.: 0.132
 Score: 82.50
 Percent Similarity: 34.13%
 Best Local Similarity: 23.35%
 Query Match: 3.55%
 DB: 5
 Gaps: 14

US-09-830-244b-2 (1-1331) x PCT-US94-14436-28 (1-375)

```

QY 380 TTACATTTTTCAGATCTTTGATGCGACTAATG-----CCATTAATGTAATGCCCTT 433
DB 83 LeuAlaPheSerAspPheLeuMetCys-LeuIleCysGlnProLeuThrValThrTyrTh 102
QY 434 ATTGGTAATACAGCATAGTTAATAAATCACTGTAACGTAATCTACACTGGATTGCTG 493
DB 102 rIleMetAspTyrTrpIlePheGlyGlnValIleuCysLysMetLeuThrPheIleGlnC 122
QY 494 CACCTCT--ACCAATAGCCTTTTGAATGACTGAAGTGTAAACAGAGAAGAGCATGT 550
DB 122 smetSerValThrValSerIleLeuSer----- 131
QY 551 CTGCAGAAAGATACCTAATATTTTGGTACTTATCTGAATCCAGATGCTGCTTC 610
DB 132 -----LeuValLeuValAlaLeuGlnArgHisGlnLeuIleLeu 145
QY 611 CCCTGAGGTGTTTCTCTTACGATCCCTCATTTGAATCCCTCGGAGCAGCAGACA 670
DB 145 nProThrGlyTrpLysProSer-----IleSerGlnAlaTyrLeuGlyI 160
QY 671 GTTAGAGAACTCTCATTTCTTTGTTTGTATTAAGACAGAGACTGCTCAAAA 730
DB 160 eValValIleTrpPheIleSer-----CysPheLeuSerLeuProPheLeuAlaAsn 178
QY 731 A-----AAGACATTTTATCATTAATACATCTTATAGCCCTAATTTCTTATCTGAA 784
DB 178 rIleLeuAsnAspLeuPheHisTyrAsn--HisSerLysValValGlnPheLeu---Gln 196
QY 785 GGCACGTGTTTTTTTAAACAGTTAAGTACTGATGTCACAGACAAATATTTCTGATC 844
DB 197 AspLysValValCysPheValSerTrpSerSerAspHisHisArgLeuIleTyr----- 214
QY 845 AGATAGTCCCTGTCACACAGTAGCAATGTGTTTCATTAAGTGGAAGAAACAGCATT 904
DB 214 ----- 214
QY 905 TTAAGTAATCTTTTGGAGACTGATTGAGTAATATAAATCTGTGCTCCTTAAGA 964
DB 214 ----- 214
QY 965 AAAAAAAAAACCTTCACCTTACGTGTGATTTATATCCCTTGAATCCCAAGTTAATTA 1024
DB 215 -----ThrThrPheLeuLeu-----PheGlnTyrCysVal 225
QY 1025 TCTTATTTCTGATATTCCTTTATATCAAGAGACCTTATCAGCCACTTCCAGA----- 1078
DB 226 ProLeuAlaPheIleLeuValCysTyrMetArgIleTyrGlnArgLeuGlnArgGlnArg 245
QY 1079 -----ACAACCCTATAGCGACCAACCATACCAACCAACATACCAACATATATAC 1126
DB 111 ----- 111

```

```

DB 246 ArgAlaPheHisThrHisThrCysSerSerArgValGlyGlnMetLysProIleAsnGly 265
QY 1127 GTTTTAATATCTCATCGTAACCTGACAGACATGATTAATGAGCTTGATGGCAATACGAC 1186
DB 266 MetLeuMetAlaMetValThrAla----- 273
QY 1187 TTC-----TACATCCATATTCCTCATCTTTGATACCATATCAGACTACTACCAC--- 1234
DB 274 PheAlaValLeuTrpLeuProLeuHisValPheAsnThrLeuGlnAspTrpTyrGlnGln 293
QY 1235 -----TTTGTGATCATCTTAAGCAATG--- 1261
DB 294 AlaIleProAlaCysHisGlyAsnLeuIlePheLeuMetCysHisLeuPheAlaMetAla 313
QY 1262 ---CGAATGTAAACCCCTATATATTTACTGATCTC 1294
DB 314 SerThrCysValAsnProPheIleTyrGlyPheLeu 325

```

Search completed: June 24, 2003, 19:26:12
 Job time : 33.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 24, 2003, 19:19:58 ; Search time 49.5 Seconds

(without alignments) 5819.122 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 2322
Sequence: 1 ccatgttttgaatcaaa.....gcttccatgacttg 1331

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 835558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus.n2p model -DEV=xip
-O=/cgn2_1/USPTO.spool/US09830244/runat.24062003.130003.10114/app_query.fasta_1.1479
-DB=Published.Applications_AA -OPM=fastest -SUPERX=n2p.rapb -MINMATCH=0.1
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pcr -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NCPU=6 -ICPU=3 -NO_MMAP -LARGESUBSTR -NEG_SCORES=0 -WAIT -DSBLOCK=100
-LONGLOC -DEV.TIMEOUT=130 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database:

Published Applications_AA:
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	6.5	78	9	US-09-924-340-6
2	150	6.5	78	9	US-09-992-600A-6
3	150	6.5	78	9	US-10-000-489-6
4	150	6.5	78	9	US-10-000-986-6

	C	5	87.5	3.8	355	9	US-10-237-563-35	Sequence 35, Appl
	C	6	85	3.7	1610	9 <td>US-10-155-933-9</td> <td>Sequence 9, Appl</td>	US-10-155-933-9	Sequence 9, Appl
	C	7	84.5	3.7	355	9 <td>US-10-237-563-27</td> <td>Sequence 27, Appl</td>	US-10-237-563-27	Sequence 27, Appl
	C	8	84.5	3.7	355	9 <td>US-10-237-563-28</td> <td>Sequence 28, Appl</td>	US-10-237-563-28	Sequence 28, Appl
	C	9	84.5	3.6	433	10 <td>US-09-841-132-363</td> <td>Sequence 363, App</td>	US-09-841-132-363	Sequence 363, App
	C	10	84	3.6	378	10 <td>US-09-971-228-7</td> <td>Sequence 7, Appl</td>	US-09-971-228-7	Sequence 7, Appl
	C	11	84	3.6	378	10 <td>US-09-842-316-7</td> <td>Sequence 7, Appl</td>	US-09-842-316-7	Sequence 7, Appl
	C	12	84	3.6	378	10 <td>US-09-731-030A-18</td> <td>Sequence 18, Appl</td>	US-09-731-030A-18	Sequence 18, Appl
	C	13	84	3.6	378	12 <td>US-10-037-610-21</td> <td>Sequence 21, Appl</td>	US-10-037-610-21	Sequence 21, Appl
	C	14	84	3.6	696	9 <td>US-09-876-790-4</td> <td>Sequence 35, Appl</td>	US-09-876-790-4	Sequence 35, Appl
	C	15	83.5	3.6	668	9 <td>US-10-011-548-35</td> <td>Sequence 29, Appl</td>	US-10-011-548-35	Sequence 29, Appl
	C	16	82.5	3.6	355	9 <td>US-10-237-563-29</td> <td>Sequence 10112, A</td>	US-10-237-563-29	Sequence 10112, A
	C	17	79.5	3.4	1329	10 <td>US-09-815-242-10112</td> <td>Sequence 15, Appl</td>	US-09-815-242-10112	Sequence 15, Appl
	C	18	79.5	3.4	2787	10 <td>US-09-893-248-15</td> <td>Sequence 32, Appl</td>	US-09-893-248-15	Sequence 32, Appl
	C	19	79	3.4	426	10 <td>US-09-841-132-573</td> <td>Sequence 362, App</td>	US-09-841-132-573	Sequence 362, App
	C	20	78.5	3.4	355	9 <td>US-10-237-563-32</td> <td>Sequence 4, Appl</td>	US-10-237-563-32	Sequence 4, Appl
	C	21	78	3.4	412	10 <td>US-09-841-132-362</td> <td>Sequence 30, Appl</td>	US-09-841-132-362	Sequence 30, Appl
	C	22	78	3.4	1332	9 <td>US-10-041-856-4</td> <td>Sequence 31, Appl</td>	US-10-041-856-4	Sequence 31, Appl
	C	23	77.5	3.4	355	9 <td>US-10-237-563-30</td> <td>Sequence 4, Appl</td>	US-10-237-563-30	Sequence 4, Appl
	C	24	77.5	3.4	355	9 <td>US-10-237-563-31</td> <td>Sequence 37, Appl</td>	US-10-237-563-31	Sequence 37, Appl
	C	25	77.5	3.4	678	9 <td>US-09-895-913A-4</td> <td>Sequence 83, Appl</td>	US-09-895-913A-4	Sequence 83, Appl
	C	26	77	3.3	355	9 <td>US-10-237-563-37</td> <td>Sequence 7, Appl</td>	US-10-237-563-37	Sequence 7, Appl
	C	27	76.5	3.3	350	9 <td>US-09-104-063-2</td> <td>Sequence 33, Appl</td>	US-09-104-063-2	Sequence 33, Appl
	C	28	76.5	3.3	350	10 <td>US-09-782-980-83</td> <td>Sequence 10, Appl</td>	US-09-782-980-83	Sequence 10, Appl
	C	29	76.5	3.3	350	10 <td>US-09-884-430-7</td> <td>Sequence 10, Appl</td>	US-09-884-430-7	Sequence 10, Appl
	C	30	76	3.3	355	9 <td>US-10-237-563-33</td> <td>Sequence 10, Appl</td>	US-10-237-563-33	Sequence 10, Appl
	C	31	76	3.3	423	10 <td>US-09-808-483-8</td> <td>Sequence 6, Appl</td>	US-09-808-483-8	Sequence 6, Appl
	C	32	76	3.3	1362	10 <td>US-09-815-242-14009</td> <td>Sequence 8, Appl</td>	US-09-815-242-14009	Sequence 8, Appl
	C	33	75	3.2	1534	9 <td>US-09-736-968A-10</td> <td>Sequence 5, Appl</td>	US-09-736-968A-10	Sequence 5, Appl
	C	34	75	3.2	1534	10 <td>US-09-736-969A-10</td> <td>Sequence 5, Appl</td>	US-09-736-969A-10	Sequence 5, Appl
	C	35	75	3.2	1534	10 <td>US-09-736-960-10</td> <td>Sequence 10, Appl</td>	US-09-736-960-10	Sequence 10, Appl
	C	36	75	3.2	2066	9 <td>US-09-978-244A-6</td> <td>Sequence 10, Appl</td>	US-09-978-244A-6	Sequence 10, Appl
	C	37	75	3.2	2107	9 <td>US-09-815-244A-6</td> <td>Sequence 10, Appl</td>	US-09-815-244A-6	Sequence 10, Appl
	C	38	74.5	3.2	621	9 <td>US-09-331-631A-5</td> <td>Sequence 1, Appl</td>	US-09-331-631A-5	Sequence 1, Appl
	C	39	74.5	3.2	495	9 <td>US-09-815-242-5331</td> <td>Sequence 12682, A</td>	US-09-815-242-5331	Sequence 12682, A
	C	40	74.5	3.2	651	10 <td>US-09-815-242-12682</td> <td>Sequence 1, Appl</td>	US-09-815-242-12682	Sequence 1, Appl
	C	41	74.5	3.2	651	10 <td>US-09-331-631A-1</td> <td>Sequence 3, Appl</td>	US-09-331-631A-1	Sequence 3, Appl
	C	42	74.5	3.2	666	9 <td>US-09-331-631A-3</td> <td>Sequence 3, Appl</td>	US-09-331-631A-3	Sequence 3, Appl
	C	43	74.5	3.2	355	9 <td>US-10-237-563-34</td> <td>Sequence 3, Appl</td>	US-10-237-563-34	Sequence 3, Appl
	C	44	74	3.2	1215	9 <td>US-10-232-539-2</td> <td>Sequence 2, Appl</td>	US-10-232-539-2	Sequence 2, Appl
	C	45	74	3.2				

ALIGNMENTS

RESULT 1
US-09-924-340-6
; Sequence 6, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: Bejani, Stephane
; APPLICANT: Tanaka, Hiroaki
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US2.REG
; CURRENT APPLICATION NUMBER: US/09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 6
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: 1..19

1264 AATGTAACCTTAATTACTGGACTCTTGGTCCAGAACTTGCCCTTTCCAAAT 1333

FILE REFERENCE: 91.US9.DIV
CURRENT APPLICATION NUMBER: US/10/000 885

FILE REFERENCE: 91.US9.DIV
CURRENT APPLICATION NUMBER: US/10/000 885

```

Length:      355
Matches:     59
Conservative: 39
Mismatch:    87
Indels:      75

```

TYPE: PRT
OCCASION: BIRTH

```

RESULT 6
US-10-155-533-9
: Sequence 9, Application US/10155533
: Publication No. US20030104003A1
: GENERAL INFORMATION:
: APPLICANT: Nguyen, Thanh V.
: TITLE OF INVENTION: A No. US20030104003A1 Surface Protein of the Malaria Parasite
: FILE REFERENCE: 4841T/CAB/R2682
: CURRENT APPLICATION NUMBER: US/10/155,533
: CURRENT FILING DATE: 2002-05-24
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 9
: LENGTH: 1610
: TYPE: PRT
: ORGANISM: Plasmodium falciparum

```


OY	875	GGTTTCATAAGTCGGAGAAACACGATTTTAAA-----GTAACCTTTTGG	921
Db	99	-----AlaLysTrpSerAspGlySerLeuLeuLeuAlaGluAspValIleAlaIleTrp	116
OY	922	GAG-----ACTGAT	930
Db	117	GluHisTrpLysGlnAlaGlyArgTyrSerLeuLeuPheGluLysLeuSerPheArgAla	136
OY	931	TTTGAGTAATAATAAACTCGTCGCCCTTACGAAAAAAACCCCTTCACCTT-----	984
Db	137	SerSerSerSerGluIleLeuIleGluIleLysGluProGluProGlnLeuAlaIle	156
OY	985	-----TACTGTGCTCATTTATATCCCTTACTGTCGCAAGTAATATCTATTT	103
Db	157	LeuAlaSerProPhePheAlaValTyrArgProGluAsnPro-----Phe	171
OY	1033	CTGGATATTTCTCT---TTTATACCAAGAGCCCTTATCAGC-----CAGTTCAG	107
Db	172	LeuSerSerGlyProPheMetProLysTrpTyrValGlnGlyInTrpLeuValLeuGln	191
OY	1078	AACAAACCACATATACGCAACACATACCAACACACAAATACCAATATACCTTTATATAC	113
Db	192	LysAsnProTyrTyrTrpAspHisAlaHisValGluLeuHisSerIleAspPheArgIle	211
OY	1138	ATCAGTAATCTGCAGGCA-----TGATTTATTCAGG-----CTTGATTCGCAATA	118
Db	212	IleProAsnIleTyrThrAlaLeuHisLeuLeuArgArgGlyAspValAspTrpValGly	231
OY	1183	CGACTTCATCAT-----CCATATTCGCATCTTCATACCAATTCACACTACACCA	123
Db	232	GlnProTyrHisGlnGlyIleProPheGlu-LeuAlaGlyTrpHisSerAlaLeuTyrThrHi	251
OY	1234	CTTTTGTGAATCATCTTAAGACCAATGCGAAATGTAAACCTATTAATTTACTGGATACT	129
Db	251	STYrProValAspGly-----ThPheTrpLeuI	261
OY	1294	CTTTCGTTCCAGA	1306
Db	261	eLeuAsnProLys	265
RESULT 10			
US-09-971-228-7			
Sequence 7, Application US/09971228			
Patent No. US2002015512A1			
GENERAL INFORMATION:			
APPLICANT: Liao, X. Charlene			
APPLICANT: Masuda, Esteban			
APPLICANT: Chu, Peter			
APPLICANT: Pardo, Jorge			
APPLICANT: Li, Congfen			
APPLICANT: Zhao, Haoran			
APPLICANT: Jlang, Yinying			
TITLE OF INVENTION: EGG: Modulators of Lymphocyte Activation and Migration			
FILE REFERENCE: 021044-000310US			
CURRENT FILING DATE: 2002-04-19			
PRIOR APPLICATION NUMBER: US/09/971,228			
PRIOR FILING DATE: 2001-04-18			
NUMBER OF SEQ ID NOS: 15			
SOFTWARE: PatentIn Ver. 2.1			
SEQ ID NO 7			
LENGTH: 378			
TYPE: PRF			
ORGANISM: Homo sapiens			
FEATURE:			
OTHER INFORMATION: human endothelial differentiation G-protein			
US-09-971-228-7			
coupled receptor (GPCR) 3 (EDG3)			

Alignment Scores:
Pred. No.: 1.37
Length: 378

Score:	84.00	Matches:	41
Percent Similarity:	42.35%	Conservative:	42
Best Local Similarity:	20.92%	Mismatches:	69
Query Match:	3.62%	Indels:	44
DB:	9	Gaps:	9

05-09-830-244B-2 (1-1331) x 05-09-971-228-7 (1-378)

Dy		65 CTCTGGGAGCAGCGACGATGTAGTAAACTCCCATTTCTTGTT-----	697
Dd		64 ILePrYsAsnSlnSySPheHisAsnAlaMeTYrPHePheIIeGLysLnLeuAlaLeu	
Oy		698 -----TTGTTTTTAAGACAGAGACTGTCTGCAAAAMAAAGCAAT	739
Dd		84 CysAsPLeuDeuAlaGLYLleAlaTYrLysValasnlIleuMeSeSLyLyLTyr	103
Oy		740 TTATCATTTAACAATCTTTATTAGACCCCTAAT-----TTCTTATCTGAAGCCACTGT	793
Dd		104 PheSerLeu-----SerProThrValTrpPheLeuArgGlnglySerMet	118
Oy		794 TTTTNTTTTAAACAGTTAGTACT-----GATGTCACACAGCAATA	835
Dd		119 PheValAlaLeuGLylLaSerThrtyrSerLeuDeuAlaIleAlaIleGLuaTrgHslLeu	138
Oy		836 TTTTCTATCATAGATAGTCCCTGTCAACAGTAGCAA-----	871
Dd		139 ThrMetIleLysMetArgProTyrgspLAsnSlnYsHtAsArgVaLrpHeLeuIle	158
Oy		872 -----TGCGGTTTCATAAAGCGGAGAACAAACACATTTTAAAGTAACTTTTGGAGA	925
Dd		159 GlyMetCystrPLeuIlleAlaPheThrIleDeuGLyLaLeu--ProIleLeuGLYTrrPasnC	178
Oy		926 CT-----GATTGAGTATATATAAAACTCGAGTCCCTCTAGGAAAAAAAACCCTTCC	979
Dd		178 ySLeuHtAsnLeuTrokspsCyseThrIleLeuProLeuTYrSerLYLs--TYrl	197
Oy		980 ACCTTTACTAGTGC--ATTATATCCCTAGTCTCCAAAGTAATATATCTTTCTCG	1030
Dd		197 LeAlaPheCySlleSerIlePheThrAlaIleLeuValTrpIleValIleLeuTYrAlaBa	217
Oy		1037 ATATTGCTTTTATACCAAGAGCCTTATCACGCCAGTTCCAGAACACCACTATACGCACA	1096
Dd		217 rGIleLyrPheLeuValLysSerSerSerAlrGLysValalaasn-----HLSA	233
Oy		1097 ACCATATCAACACCAATACCAACAATATAGCTTTTATATATACAC	1140
Dd		233 sTsAsnSerGIuArgSerMetAlaLeuLeuArGrHrValValIle	247

RESULT 11
 US-09-842-316-7
 Sequence 7, Application US/09842316
 Patent No. US20020099191A1
 GENERAL INFORMATION:
 APPLICANT: KOSTENSIS, Eva
 APPLICANT: GASSENHUBER, Johann
 TITLE OF INVENTION: ED58 RECEPTOR, ITS PREPARATION AND USE
 FILE REFERENCE: 38005-147
 CURRENT APPLICATION NUMBER: US/09/842.316
 CURRENT FILING DATE: 2001-04-26
 PRIOR APPLICATION NUMBER: EP 116589.3
 PRIOR FILING DATE: 2000-08-01
 PRIOR APPLICATION NUMBER: EP 108858.2
 PRIOR FILING DATE: 2000-04-26
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: Patentln version 3.1
 SEQ ID NO 7
 LENGTH: 378
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-842-316-7

Alignment Scores: .
Pred. No.: 1.37
Length: 378

Score: 84.00
Percent Similarity: 42.358
Best Local Similarity: 20.928
Query Match: 3.628
DB: 10
Matches: 41
Conservative: 42
Mismatch: 69
Indels: 44
Gaps: 9

US-09-830-244b-2 (1-1331) x US-09-842-316-7 (1-378)

```

OY 653 CTCGCGACACAGACAGACAGTGTAGACCTCCATCTTCTTGT----- 697
DB 64 ILETPLYSASnslsnphehlsasnaargmettyrphelleglyasnualeu 83
OY 698 -----TTGTTTTTAAACAGACAGCTGTCTCAAAAAAGACAT 739
DB 84 CysAspLeuLeuAlaGlyLeuAlaTyLysValAsnIleLeuMetSerGlyLysThr 103
OY 740 TTATCATTTAAACATCTTTAGAGCCCTTAAT-----TTCTTATCTGAAGCAGCTGT 793
DB 104 PheSerLeu-----SerProthrValITrPheLeuArgGluGlySerMet 118
OY 794 TTTTCTTTTAAACAGTAACTACT-----GATGTCAACAGACAATA 835
DB 119 PheValAlaLeuGlyAlaSerThrCysSerLeuLeuAlaIleAlaIleGluArgHisLeu 138
OY 836 TTTGATCAGATAGTCCCTGTCAACAGTACAA----- 871
DB 139 ThrMetIleLysMetArgProtyrAspAlaAsnLysArgHisArgValPheLeuLeuIle 158
OY 872 -----TGCGTTTCATTAAGTGGAGAGAAACAGCATTTAAAGTACTTTTGGGAGA 925
DB 159 GlyMetCysTrpLeuIleAlaPheThrLeuGlyAlaLeu--ProIleLeuGlyTrpAsn 178
OY 926 CT-----GATTGATGATATATAAAGCTGTGTCTCCCTTAAGAAAAAACCCCTCC 979
DB 178 YsLeuHisAsnLeuProAspCysSerThrIleLeuProLeuTySerLysLys--TyrI 197
OY 980 ACCTTACTGTGC---ATTATATCCCTTAGTCCAAAGTAAATATCTTATTTCTGG 1036
DB 197 leaIaPheCysIleSerIlePheThrAlaIleLeuValIThrIleValIleLeuTyAla 217
OY 1037 ATATTGCTTTTATACCAAGAGCTTTATCAGCCAGTCCAGAACACACTATATCCACA 1096
DB 217 rglIeTyRphLeuValLysSerSerArgLysValAlaAsn-----HisA 233
OY 1097 ACCATACCAACCAATACCAACAATATACGTTTATATATCATC 1140
DB 233 snasnsrGluArgSerMetAlaLeuLeuArgThrValValIle 247

```

RESULT 12
US-09-731-030A-18
Sequence 18, Application US/09731030A
Patent No. US20020142375A1
GENERAL INFORMATION:
APPLICANT: MONROE, Donald G.
APPLICANT: GUPTA, Ashwani K.
APPLICANT: ZASTAMNY, Roman L.
TITLE OF INVENTION: MAMMALIAN EDG-7 RECEPTOR HOMOLOGS
FILE REFERENCE: 8074-0015
CURRENT APPLICATION NUMBER: US/09/731,030A
CURRENT FILING DATE: 1998-12-29
PRIOR APPLICATION NUMBER: 60/070,184
PRIOR FILING DATE: 1997-12-30
NUMBER OF SEQ ID NOS: 21
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 18
LENGTH: 378
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: Mammalian EDG
US-09-731-030A-18

Alignment Scores:

Pred. No.: 1.37
Score: 84.00
Percent Similarity: 42.358
Best Local Similarity: 20.928
Query Match: 3.628
DB: 10
Length: 378
Matches: 41
Conservative: 42
Mismatch: 69
Indels: 44
Gaps: 9

US-09-830-244b-2 (1-1331) x US-09-731-030A-18 (1-378)

```

OY 653 CTCGCGACACAGACAGACAGTGTAGACCTCCATCTTCTTGT----- 697
DB 64 ILETPLYSASnslsnphehlsasnaargmettyrphelleglyasnualeu 83
OY 698 -----TTGTTTTTAAACAGACAGCTGTCTCAAAAAAGACAT 739
DB 84 CysAspLeuLeuAlaGlyLeuAlaTyLysValAsnIleLeuMetSerGlyLysThr 103
OY 740 TTATCATTTAAACATCTTTAGAGCCCTTAAT-----TTCTTATCTGAAGCAGCTGT 793
DB 104 PheSerLeu-----SerProthrValITrPheLeuArgGluGlySerMet 118
OY 794 TTTTCTTTTAAACAGTAACTACT-----GATGTCAACAGACAATA 835
DB 119 PheValAlaLeuGlyAlaSerThrCysSerLeuLeuAlaIleAlaIleGluArgHisLeu 138
OY 836 TTTGATCAGATAGTCCCTGTCAACAGTACAA----- 871
DB 139 ThrMetIleLysMetArgProtyrAspAlaAsnLysArgHisArgValPheLeuLeuIle 158
OY 872 -----TGCGTTTCATTAAGTGGAGAGAAACAGCATTTAAAGTACTTTTGGGAGA 925
DB 159 GlyMetCysTrpLeuIleAlaPheThrLeuGlyAlaLeu--ProIleLeuGlyTrpAsn 178
OY 926 CT-----GATTGATGATATATAAAGCTGTGTCTCCCTTAAGAAAAAACCCCTCC 979
DB 178 YsLeuHisAsnLeuProAspCysSerThrIleLeuProLeuTySerLysLys--TyrI 197
OY 980 ACCTTACTGTGC---ATTATATCCCTTAGTCCAAAGTAAATATCTTATTTCTGG 1036
DB 197 leaIaPheCysIleSerIlePheThrAlaIleLeuValIThrIleValIleLeuTyAla 217
OY 1037 ATATTGCTTTTATACCAAGAGCTTTATCAGCCAGTCCAGAACACACTATATCCACA 1096
DB 217 rglIeTyRphLeuValLysSerSerArgLysValAlaAsn-----HisA 233
OY 1097 ACCATACCAACCAATACCAACAATATACGTTTATATATCATC 1140
DB 233 snasnsrGluArgSerMetAlaLeuLeuArgThrValValIle 247

```

RESULT 13
US-10-037-616-21
Sequence 21, Application US/10037616
Patent No. US20020123148A1
GENERAL INFORMATION:
APPLICANT: English, Denis
APPLICANT: Kovacs, Richard J.
APPLICANT: Rizzo, Maria T.
APPLICANT: Silva, Daniel T.
TITLE OF INVENTION: Sphingolipid Compositions and Methods for Their Therapeutic Use
FILE REFERENCE: 7042-119
CURRENT APPLICATION NUMBER: US/10/037,616
CURRENT FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: 60/243,887
PRIOR FILING DATE: 2000-10-27
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn version 3.1
SEQ ID NO 21
LENGTH: 378
TYPE: PRT
ORGANISM: Homo sapiens
US-10-037-616-21

Alignment Scores:

/ 2 Gruppenserie-----

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 24, 2003, 19:16:52 ; Search time 44 Seconds

(without alignments)
5816.133 Million cell updates/sec

Title: US-09-830-244b-2

Sequence: 1 ccatgttttagatcaacaag.....gcctttccaatgtcacttg 1331

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPRO-SPool/US09830244/unal_24062003_130002_10054/app_query.fasta.1.1479
-DB=PIR.73 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-DOCLIN=200 -THR=SCORE=1 -END=1 -MATRIX=DISSUM62 -TRANS=human40.cdi -LIST=45
-OUTFMT=ptc -NORTEXT -HEAPSIZE=500 -MINLEN=0 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-USER=US09830244.GCGN.1.1.71 -runat_24062003_130002_10054 -NCPU=6 -ICPU=3
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELop=6 -DELext=7

Database:

PIR.73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	136	5.9	62	1	SBHUP
2	91.5	4.0	949	2	F90086
3	91.5	4.0	1138	2	S64484
4	91	3.9	2708	2	T09079
5	91	3.9	2819	2	T09080
6	90	3.9	691	2	T46476
7	89	3.8	445	2	A44848
8	88.5	3.8	474	2	T31104
9	88	3.8	1005	2	F90099
10	87.5	3.8	332	2	T25023
11	87	3.8	855	2	T17460
12	86	3.7	2028	2	T08025
13	85.5	3.7	360	2	A53611
14	85.5	3.7	1070	2	G84982

15	85	3.7	61	1	SBHUP	stathurin precursor
16	85	3.7	567	2	S58750	NADH2 dehydrogenas
17	85	3.7	577	2	T04229	ABC-type transport
18	85	3.7	580	2	T18439	hypothetical prote
19	85	3.7	600	2	T18446	hypothetical prote
20	85	3.7	620	2	A58932	cytochrome C-type
21	85	3.7	2391	2	T18410	cardamoyl-phosphat
22	84	3.6	378	2	UC5245	G protein-coupled
23	84	3.6	651	2	F90111	DNA primase [impor
24	84	3.6	748	2	S54505	hypothetical prote
25	83.5	3.6	331	2	T26437	ATP-dependent RNA
26	83.5	3.6	515	2	F90366	tubulin beta chain
27	83	3.6	442	1	UBHUP	nuclear pore prote
28	83	3.6	1037	2	S73879	protein g377 - mal
29	83	3.6	3119	2	T37879	ERD1 protein - yen
30	82.5	3.6	362	2	T18414	hypothetical prote
31	82.5	3.6	475	2	S69698	replication factor
32	82.5	3.6	1847	2	G59105	hypothetical prote
33	82	3.5	339	2	D60406	interleukin-8 rece
34	82	3.5	356	2	S42096	hypothetical prote
35	81.5	3.5	383	2	E82897	lipoprotein [impor
36	81	3.5	368	2	C90558	conserved hypotec
37	81	3.5	835	2	F64246	hypothetical prote
38	80.5	3.5	461	2	H90090	beta-tubulin 1 - a
39	80	3.4	445	2	J00422	MG148 homolog VXP
40	80	3.4	445	2	S73996	DNA-directed RNA p
41	80	3.4	960	2	S72284	hypothetical prote
42	80	3.4	1186	2	S70430	hypothetical prote
43	79.5	3.4	365	2	E90099	hypothetical prote
44	79.5	3.4	376	2	T24368	hypothetical prote
45	79.5	3.4	446	2	H90094	hypothetical prote

ALIGNMENTS

RESULT 1

SBHUP

stathurin precursor [validated] - human

C:Species: Homo sapiens (man)

C:Date: 24-Apr-1984 #sequence_revision 30-Jun-1993 #text_change 08-Dec-2000

C:Accession: JH0153; A27308; B27489; A03288; A32524

R:Sablatni, L.M.; He, T.Z.; Azen, E.A.

Gene 89, 245-251, 1990

A:Title: Structure and sequence determination of the gene encoding human salivary sea

A:Reference number: JH0153; MUID:90323623; PMID:2373365

A:Accession: JH0153

A:Molecule type: DNA

A:Residues: 1-62 <SA2>

A:Cross-references: GB:M31077

R:Sablatni, L.M.; Carroll, L.R.; Johnson, G.W.; Azen, E.A.

Am. J. Hum. Genet. 41, 1048-1060, 1987

A:Title: cDNA Cloning and chromosomal localization (4q11-13) of a gene for stathurin,

A:Reference number: A27308; MUID:88074310; PMID:3502720

A:Accession: A27308

A:Molecule type: mRNA

A:Residues: 1-62 <SAB>

A:Cross-references: GB:M32639; NID:g338504; PIDN:AAA60593.1; PID:g338506

R:DiChinson, D.P.; Ridall, A.L.; Levine, M.J.

Biochem. Biophys. Res. Commun. 149, 784-790, 1987

A:Title: Human submandibular gland stathurin and basic histidine-rich peptide are enc

A:Reference number: A27489; MUID:88106506; PMID:3426601

A:Accession: B27489

A:Molecule type: mRNA

A:Residues: 1-62 <DIC>

A:Cross-references: GB:M18371; NID:g338610; PIDN:AAA60600.1; PID:g338611

R:Schlesinger, D.H.; Hay, D.I.

J. Biol. Chem. 252, 1688-1695, 1977

A:Title: Complete covalent structure of stathurin, a tyrosine-rich acidic peptide whi

A:Reference number: A03288; MUID:77118656; PMID:838735

A:Accession: A03288

A:Molecule type: protein

A:Residues: 20-62 <SCH>

R:Oppenheim, F.G.; Hay, D.I.; Smith, D.J.; Offner, G.D.; Troxler, R.F.

RESULT 5

F:Species: Plasmodium falciparum
 C:Species: Plasmodium falciparum
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 21-Jul-2000
 C:Accession: T09080
 R:Su, X.Z.; Kirkman, L.A.; Fujitaka, H.; Wellem, T.E.
 Cell 91, 593-603, 1997

A:Reference number: 216556; MUID:98054002; PMID:9393833
A:Accession: T09080
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2819 <SUXX>
A:Cross-references: EMBL:AF030693; NID:q2642515; PIDN:AA047854.1; PID:q2642516
A:Experimental source: strain HB3; from Honduras
C:Genetics:
A:Gene: c92
C:Keywords: toxin resistance

Alignment Scores:

Record No.:	0.569	281
Score:	91.00	41
Percent Similarity:	43.85%	Matches:
Best Local Similarity:	21.93%	Conservative:
Query Match:	3.92%	Mismatches:
DB:	2	Indels:
		Gaps:

OS-09-830-244B-2 (1-1331) x T09080 (1-2819)

```

0Y      669 CAGTAGAGAGAACCTCCATTCCTTCTGTTTGGTTTTTTTAAACAGACAGACCTCTGTCCAA 728
      ||| ||| ||| |||
Db      580 GlnValPheMetLeuProPheSerTyrPheSerAsnLysLysLysLysLysLysLys 728
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y      729 AAAAAGCATTATATACATTATACATCTTATTAGAGCCCAATTTCTATCTGACG--- 785
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      597 LysLys-----AsnIleIlehelLysAsnIleLeuSerTyrTyrGluTyr 611
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y      786 GCACGCTTTTTTTTTTAAACAGATTAACTA-----CTGATGTCACACAGCAAAATTTTC 839
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      612 ValIleCysPhePheMetLysLysLysIleLysAsnIleLeuSerTyrAsnProTyrTyr 631
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y      840 TGATCAGATATCCCTCTGTCAACAGTAGCAGCAATCTGGTTCATTAAGTGGAGAAACAA 899
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      632 -IleHisIle-----GluSerTyrPyrAsnLysIleOly----- 642
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y      900 GCATTTTAAAGTAACTTTTGGGAGACTGATTG-----AGTAAATA 941
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      643 ----TyrCysIleValPhePheLeuPheSerIleLeuLysIlePheCysThrCysAspAs 661
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y      942 TAAACCTGGTCTCCCTTAAGAAAAAACCTTCACACTT----- 984
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      661 PAspThrThrLeuLeuIleSerArgLysHisValTyrTyrLeuSerAspSerLeuGlu 681
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y      985 -----TACTGTCTCATTTATATCCCTTACCTTCAAGTTCATTAATCTTATT 1031
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      681 sSerGluTyrSerTyrGluThrGlyValIleAsnIleAsnLeuAspLeuLeuAspTyr 701
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y      1032 TCTGGATTTGCGTTTTATACCAAGACGCTTATACACCGA----- 1071
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      701 ePheLeuPheLeuHisPheIleLysAspGluLeuLysCysGlnAspThrAsnGluAsnIle 721
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y      1072 -----TTCCAGAACACACCATATACGCACACACATACCAACACCAATAT 1124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      721 sPheAspGlnMetLeuAsnHisValLysAsnAspHisThrAsnTyrTyrGlnAsnAspAs 741
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y      1125 ACCTTTTAAATCATCACT 1143
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 741 nilepheAsnLeuLeuThr 747

RESULT 6
T46476

hypothetical protein DKFZp340c0921.1 - human
C/Species: Homo sapiens (man)
C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C/Accession: T46476
R/Baum: H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A/Reference number: Z23034

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-691 <AA>
A:Cross-references: EMBL:ALJ37718
A:Experimental source: adult testis, clone DKE2p434C0931
C:Genetics:
A:Note: DKE2p434C0931.1

Alignment Scores:

Pired. No.:	0.624	length:	69
Score:	90.00	Matches:	46
Percent Similarity:	40.708	Conservative:	35
Best Local Similarity:	23.128	Mismatches:	63
Query Match:	3.898	Indels:	55
DB:	2	Gaps:	9

US-09-830-244B-2 (1-1331) x T46476 (1-691)

[illegible]

C:Species: Physarum polycephalum
 C>Date: 31-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 23-Feb-1997
 C:Accession: A44848
 J:Paul, E.C.; Buchschacher Jr., G.L.; Cunningham, D.B.; Dove, W.F.; Burland, T.G.
 J:Gen. Microbiol. 138, 229-238, 1992
 A:Title: Preferential expression of one beta-tubulin gene during flagellate development
 A:Reference number: A44848; MID:92211323; PMID:1556551
 A:Contents: amoebae, flagellates
 A:Accession: A44848
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-445 <PAU>
 A:Note: sequence extracted from NCBI backbone (NCBIN:92291, NCBI:92292)
 C:Superfamily: tubulin
 C:Keywords: nucleus

Alignment Scores:
 Pred. No.: 0.761 Length: 445
 Score: 89.00 Matches: 44
 Percent Similarity: 39.81% Conservative: 38
 Best Local Similarity: 21.36% Mismatches: 76
 Query Match: 3.83% Indels: 48
 DB: 2 Gaps: 8

US-09-830-244b-2 (1-1331) x A44848 (1-445)

OY 386 TTTTCAGATCTTTGATGTCGACATAATGTAATGCGCTATTGTAATAC 445
 DB 159 TTTTCAGATCTTTGATGTCGACATAATGTAATGCGCTATTGTAATAC 445
 OY 446 AGCATAGTAAATTAACGTTACGTAATGTAATGCGCTATTGTAATAC 481
 DB 179 ValValGluProTyrAsnAlaThrLeuSerValHisGlnLeuValGluAsnAlaAspGlu 198
 OY 482 -----TTGATTTGCTGACCTTACCAATAGCCTTTGAATGACGAAAGTGA 532
 DB 199 ValMetCysIleAspAsnGluAlaLeuTyrAspIleSerPheArgThrLeuThr 218
 OY 533 ACAGAGAAAGAGCATCTGTCAGAAAGAGATGTAATTTTGGTACTTATCTGA 592
 DB 219 ThrProThrTyrGlyAspAsnHisLeuValSerAlaVal-----MetSer 234
 OY 593 AATCCAGATGCTGCTGCTGCGAGACAGACAGATGTAAGAACTCCATTTCTTGT 640
 DB 235 GlyIleThrCysCysLeuArgPheProGlyGlnLeuAsnSerAspLeuArgLysLeuAla 254
 OY 641 ----TCAATTGAATCCCTCTGGGAGACAGACAGATGTAAGAACTCCATTTCTTGT 697
 DB 255 ValAsnLeuIleProPhe-----ProArgLeuHisPhePheLeu 267
 OY 698 TTGTTTAAAGACAGACTCTGTCATAAAAAAGACATTTATCATTA----- 748
 DB 268 ValGlyPheAlaProLeuThrSerArgLysValGlyTyrArgSerLeuThrValPro 287
 OY 749 -----TAACTCTTATTAGACCCCTAAT----- 772
 DB 288 GluLeuThrGlnGluMetPheAspAlaLysAsnMetAlaAlaSerAspProArgHis 307
 OY 773 -----TTCTATCTGAGAGCAGCTCTTTTAAACAGTAAAGTACAGATGTCAC 826
 DB 308 GlyArgTyrLeuThrAlaSerAlaMetPheArgLysArgMetSerThrLysGlnValAla 327
 OY 827 AGACAAATATTTCTGATAGTATGTCCTGTCACAAAGTACCAATGCTTTCATAAG 886
 DB 328 GluGlnMetLeuAsnValGln-----AsnLysAsnSerSerTyrPheValGlu 343
 OY 887 TGGAGAGAAACAGCATT 904
 DB 344 TrpIleProAsnAsnIle 349

RESULT 8
 T31104
 hemolysin accessory protein homolog - Haemophilus ducreyi (fragment)

C:Species: Haemophilus ducreyi
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
 C:Accession: T31104
 J:Ward, C.K.; Lumley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
 J:Bacteriol. 180, 6013-6022, 1998
 A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
 A:Reference number: 220984; MID:99030326; PMID:9811662
 A:Accession: T31104
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-474 <MAR>
 A:Cross-references: EMBL:AF057696; MID:93929021; PID:93929022; PIDN:AACT9760.1
 A:Genetics:

Alignment Scores:
 Pred. No.: 0.867 Length: 474
 Score: 88.50 Matches: 42
 Percent Similarity: 38.46% Conservative: 33
 Best Local Similarity: 21.54% Mismatches: 71
 Query Match: 3.83% Indels: 49
 DB: 2 Gaps: 10

US-09-830-244b-2 (1-1331) x T31104 (1-474)

OY 885 TTTATGAACACATTTGCTACTGTTGACAGGAGC-----TATCTGATCAGAAATTTTGT 829
 DB 82 PheValIleProAsn-----LeuSerGlyLysIlePheSerIleHisAspLeuAsp 98
 OY 828 CTGTGACATCAGTACTTAATCTGTTTAAAAAAAACA-----GTGCTTTCAGAT 778
 DB 99 GlnLeuValGluValLeuAsnThrValAsnLysArgAlaGluIleLysValLeuAlaSer 118
 OY 777 AAGAAATAGAGGGCTCTAATAGATGTTATATGATAATGCTTTTGGAGACA 718
 DB 119 LysAlaTyrGlyLys-----SerAsn 125
 OY 717 GTCTGTGCTTAAAAAACAACAAGAAATGAGAGTTCATCTGCTGCTGCTC 658
 DB 126 LeuAsnIleLeuThrGluArgThrArgLysTrpProThr-----ValThrLeuSerIleAsn 144
 OY 657 CAGAGGGATTCATGAGATCGTACAGAGAAACACCTGACAGGAGACAGATCTT 598
 DB 145 AsnSerGly-----LysGlnAsnAsnGlnAsnGlnArgAsnGlnMet 158
 OY 597 GGATTCAGATTAAGTACCAAAAAATATATGCTATCTTTTCGACAGATGCTTTC 538
 DB 159 ThrLeuAsnValSerTrpSerAspLeuGlyThrAsn----- 171
 OY 537 TCTGTAACTTTCACCTTCACTTCAAAAAGCTATTGGTAGAGTGCAGCAATCCAGTGT 478
 DB 172 -----AspValIlePheSerPheLysThrGlyTyr-----ArgLeuTyrLysGluThrLys 187
 OY 477 AGATTACTTACACAGTTATTTTAACTATGCTTATTCACCAATTAAGCATTACCAATA 418
 DB 188 LysAsnThrGlnGlnAsnTyrSerLeuSerTyrIleGlnProPheSerTyrTyrThrLeu 207
 OY 417 ATGCAATTAAGTGCACATCAAAAAGATCTGAAATGCTAATAGGG----- 376
 DB 208 -----GluLeuLysAlaSerGlnSerAlaTyrAsnLysGluLeuThrGlyPhe 223
 OY 375 ----ACATTTCCCAACAAGAAAGAAATCCATGCTCTTATAA 334
 DB 224 TyrThrTyrProSerSerGlyLysThrGlnThrAlaAsnIleLys 238

RESULT 9
 F90099
 hypothetical protein orf1005 (imported) - Guillardia theta nucleomorph
 C:Species: nucleomorph Guillardia theta
 A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
 C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
 C:Accession: F90099
 R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.;

Nature 410, 1091-1096, 2001
 A:Title: The highly reduced genome of an enslaved algal nucleus.
 A:Reference number: A99082; MUID:11332671; PMID:11332671
 A:Accession: F90099
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1005 <DO>
 A:Cross-references: GB:AF165818; NID:g13794550; PIDN:AAK39925.1; GSPDB:GN00150
 C:Genetics:
 A:Gene: orf1005
 A:Map position: 1
 A:Genome: nucleomorph
 C:Keywords: nucleomorph

Alignment Scores:
 Pred. No.: 1.07 Length: 1005
 Score: 88.00 Matches: 99
 Percent Similarity: 37.21% Conservative: 45
 Best Local Similarity: 25.38% Mismatches: 125
 Query Match: 3.81% Indels: 118
 DB: 2 Gaps: 22

US-09-830-244B-2 (1-1331) x F90099 (1-1005)

```

OY 999 AATAATGACACGATGAAGGTGAGAGGTTTCTTAAGGAGACAGGATTTAT 940
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 132 llaasnastrhrvalserleuylsasnleuphetyr--LysCysaspasnleupheile 150
OY 939 ATTACTCAA-----
   :::::
Db 151 leuilegilyltyrserglycysargthrleuylsasnleupheylsasnleuargleu 170
OY 930 -----ATCAGTCTCCCAAAAAGTACTTAAATGCTG 898
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 171 leuglnpheasnleuylsasnasthrilleasntrlyrtpylslystyr---LysSerille 189
OY 897 TTTTCT-----TCCCACTTATGAACACACATTTGCTACTGTTGACAGG 853
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 190 PheSerlyLysLysLysPheLysLysPheGlnLysGlnLysLysLysLysLysLysLys 206
OY 852 GACTATCTGATCAGAAATATTGCTGCTGCTGATCAGTACTGATCTTAAATGCTGCT 793
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 207 AspArglyrilearg-----LeupheSerilleleuaspasnlyrlyls----- 220
OY 792 ACACTGCTTCAGTAAAGAAATTTGGGCTCTAATTAAGATGATATGATTAATGCTCT 733
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 221 -----lleaspyrlylslsleuphetyrillephehistrlyrtpglnthrlyls 237
OY 732 TTTTGTGAGACAGAGCTCTGCTCTTAAACAAAACAAGAAATGGAGAGTCTACTA 673
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 238 lleleu-----LysLysAsnilleargGlnLysLysaspCysAsnleuValilleleu 255
OY 672 ACTGTCTCTGCTCCAGAG-----
   :::::
Db 256 LyslleasnleuphethrlylspheGlnLysLysAsnastyrserasntrlyrlystrille 275
OY 651 -----GGATTCAATGAGATCTGTAAGAGAAACACCTCGAGGGAGACATCTT 598
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 276 PheTyrrPhePheasnserasnLysLyslleasnAsnillelleLysasnleuGlnphe 295
OY 597 GGATTTCAGATTAAGTACCAAAAATATTAGTACTCTTCTTGCGACGATCCCTCTTC 538
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 296 Asnphse--AspLysleu-----AsnSerlystrileuphe----- 306
OY 537 TCTGTTAACACTTTCAGTACTTCAAAAGCTATTGGTAGAGGTGACGAATCAAGTGT 478
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 307 -----SerlySerleuaspLeuargProlyasn----- 316
OY 477 AGATTACTGTACAGATTATTAATAGTCTGTATTCACCAATACGCAATACCATTA 418
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 317 -----GluSerilleMetAsnMetillePheValThrGlnSerasnilleLysAsna 333
OY 417 ATGGCATTTAGTGACATCAAAAAGATCTGMAATGCTTAATGGACATTTCCACAAAAGA 358
  
```

```

Db 333 snValalaValylsargilleLysValyls-----LysLysSerl 347
OY 357 AAGAAATCCAAATGGCTCTCTTAAGGAGCAATATACCAATGCGACAGATGAATTT 298
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 347 yslleuiletyrlylserLysGlnlleSerillethrhrPheLysTyrrPheLysllel 367
OY 297 CTGTCAAAAGAGTACTCTTCAACCAATTCAGAAACATGTTCCTCAATTAAGCCCTGG 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 367 lePheLysLysLysTyrrPhe-----GlnAsnlyslleuLeuSerleupheilel 384
OY 237 GGAGCCACATTAGC---CACTATCTCAGATACTGATTTATTTGTT-----AGCCT 190
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 384 leuMetilleSerlyshistyrille-AspLys-ileValleuTyrrLysLysAsnSerle 403
OY 189 TAGGGGCGCCCAAT-----ATTCGCCCTCAT-----CTGCTAACTTAACCTT 148
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 403 uleuGlyTyrrAsnillellelleasnGlnhisGlyValargGlyValleuAsn---Asnle 422
OY 147 GTTTACTCACAAGTCTGATTAATTCAGCCAGTACTCTT-----AACATTCC 100
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 422 uSerlyProLeuAspSerlystyrargasnserasniletyraspLysCysHisilleVa 442
OY 99 CTATGGCAGTGCAGAAATTTCCACAGATTTGTAACACTGTAATTTACGTTTGTAATC 40
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 442 lPhePheSerArglystyrPheThrillePheLys-----AsnTyrrPhePhe----- 457
OY 39 AGCAGCTTACCG 27
   |||||:|||||
Db 458 ----Serleupro 460
  
```

RESULT 10
 T25023
 Hypothetical protein T20B3.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T25023
 R:Motifmore, B
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19970
 A:Accession: T25023
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-332 <MW>
 A:Cross-references: EMBL:Z81593; PIDN:CA804743.1; GSPDB:GN00023; CESP:T20B3.5
 A:Experimental source: clone T20B3
 A:Gene: CESP:T20B3.5
 A:Map position: 5
 A:Introns: 110/1; 246/3
 C:Superfamily: Caenorhabditis hypothetical protein C4967.2

Alignment Scores:
 Pred. No.: 1.07 Length: 332
 Score: 87.50 Matches: 64
 Percent Similarity: 33.99% Conservative: 40
 Best Local Similarity: 20.92% Mismatches: 87
 Query Match: 3.77% Indels: 115
 DB: 2 Gaps: 15

US-09-830-244B-2 (1-1331) x T25023 (1-332)

```

OY 529 GTTACAGAGAAAGAGCATGCTCGCAAGAAAGATAGCTAATTTTGTG----- 579
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 89 ValLeuLysGlnleGlyValSerValGlnAlaGlnVal--TyrrPheValValleuVal 107
OY 580 -----
   :::::
Db 108 GlyAlaTyrrMetMetValAlaIleSerilleValPheGlnAsnArgLeupheValleuVal 127
OY 589 CTCGAATCCCAAGATGCTG-----CTTCCCTGCAGGTGTT----- 624
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 128 MetThrAsnLysMetLeuHisLysPheAlaLeuProIleTyrrilleHisTyrrillePhe 147
  
```

Score:	87.00	Matches:	71
Percent Similarity:	34.06%	Conservative:	38
Best Local Similarity:	22.19%	Mismatches:	116
Query Match:	3.76%	Indels:	95
DB:	2	Gaps:	15
US-09-830-244B-2 (1-1331) x T17460 (1-855)			
QY	906 AAAATGCTTTTTCCTCCACCTTATGAAACACATTTGCTACCTGTGACAGGGACATAT	84	
Db	5 LysLeuValPhePheMetArgTyrVal -----PheLeuPheGlnLeuValSerSerSer	22	
QY	846 CTGATGAGAAATATTTTCTGTGACATACACTACTTAACTGTTAAAAAAAACAGTG	78	
Db	23 SerLeuAlaGlnHisLeuGlyS -----	28	
QY	786 CCTTCAGATTAAGAAATTAGGGGCTCTAATTAAGATGTTATAT	74	
Db	29 ProGluAspGlnAlaLeu -----AlaLeuLeuGlnPheLeuAsnMetPheIleValAsnPro	47	
QY	744 -----GATAAATGTCCTTTTTTTGGACACAGAGCTGTGCTTAAAAACAACA	69	
Db	48 AsnAlaPheHisTyrCysProAspIlePheGlyArgTyrGluIleGlnSerTyrProArgThr	67	
QY	693 AAGAAATGAGAGATTACTACTGCTCTGTCGCCAGAGGAGATTCATGAGAGATCGT	63	
Db	68 LeuSerIlePheAsn -----LysSerThrSerCysCysSerIlePheArgIleValHisCysAspGlu	86	
QY	633 AAGAAG-----GAAACACACTGACGGGGAAGACAGATCTTGA--	59	
Db	87 ThrIleArgIleValIleAlaLeuAspLeuGlnLeuIndIlePheHisSerAsnSer	106	
QY	594 -----TTTCAGATTAAGATTCACAAAATAATTTAGCTATCTTTCTGACAGATGCTCT	54	
Db	107 SerLeuPheGlnLeuSerAsnLeuValArgLeuAspLeuSerPhe-----	121	
QY	540 TTCTCTGTAAACACTTTCAGTCACTTCAAAAAGCTATTGGTAGAGGTGACACAAATCCAG	48	
Db	122 -----AsnAspPheIleGlySerPro-----PheSerProIlePhe	133	
QY	480 TGTAGATTACTGTAACAGCTTTTATTAAGTCTGTATTCACAAATAGCGGATTTACA	42	
Db	134 GlyIlePheSerAspLeuThrIleHisLeuAspLeuSerHisSerSerPheIleGlyIle	153	
QY	420 ATATGCACTTAATGGACATCAAAACAAATGTCGAATAGCGATATTTCCACAAA	36	
Db	154 ProPheGlnIleSerHisLeuSerIleValIleValArg-----	167	
QY	360 GGAAGAAATCCCAATGGCTCTATAAAGAGAAATATAGCATGTGGCAGACAGATGAAA	30	
Db	168 -----IleArgGlyIleIleValIleLeuSerLeuValProHisAs	180	
QY	300 TTTCCTCAAAAGAAAGTACTTCTCAAGCAATCAGAAAACCTTCCCAATTAAGCC	24	
Db	180 nPhe-----GluLeuLeuLeuValAsnLeuIlePheGlnLeuArg--	192	
QY	240 TGGGAGCCACATATGACCATATCTCAGATACATGATATTTGTTTACCTTAAGGGCT	18	
Db	193 -----AspLeuGlnLeuGlnIleAsnIleSerSerIleValIleProse	207	
QY	180 CAAATATCCCTCATCTGCTAAATCTCAACATG	14	
Db	207 rAsnPheSerSerHisLeuThrAsnLeuArgLeuProPheThrGluIleValArgIleLe	22	
QY	146 -----TTTACTCAACGTGTGAT-----AAATTCACAGCCAG	11	
Db	227 nProGluArgPheHisIleSerAsnLeuGlnIleValAspLeuSerPheAsnPro--	246	

```

Db      247 -GlnLeuThrValAlaArgPheProThrThrLysSTyrosnSerSerAlaSerLeuValasn 265
RESULT 12

```

DNA-directed RNA polymerase beta' chain homolog - Chlamydomonas reinhardtii chloroplast
708025
M:Alternate names: rPOC2 protein
C:Species: Chloroplast Chlamydomonas reinhardtii
I:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 08-Oct-1999
R:Nuotio, S.; Purton, S.
submitted to the EMBL Data Library, May 1996
A:Description: The chloroplast rPOC2 gene of Chlamydomonas reinhardtii.
A:Reference number: Z16398
A:Accession: J08025
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2028 <NDO>
C:Cross-references: EMBL:U57326; NID:q1354831; PIDN:AAB0197.1; PID:q1354832
C:Genetics:
A:Gene: rPOC2
A:Genome: chloroplast
C:Keywords: chloroplast

Alignment Scores:
Pred. No.: 1.89 Length: 2028
Score: 86.00 Matches: 37
Percent Similarity: 44.87% Conservative: 33
Best Local Similarity: 23.72% Mismatches: 62
Query Match: 3.70% Indels: 24
DB: 2 Gaps: 8

US-09-830-244B-2 (1-1331) x J08025 (1-2028)

QY 835 ATTTCGATCAGATAGTCCCTGTCAACAG--TAGCAATGTGCTTCATTAAGTGGCA 891
||| ||||: ||: |||||: ||| ||| |||
Db 1363 ILELeuApLysThrIleHisCysGlnLysProThrLysValLeuPheLysSeryl 1382
QY 892 -----AGAAACAGCATTTTAAAGTAACTTTTGGAGACTGATTTGAGTAATATAA 945
:::|||||||: ||: ||| |||
Db 1383 PheSerLysLysGlnHisIleTyrTyrLeuGluPheLeuAsnThr-----LysAsnHisArg 1400
QY 946 ACTGTGCTGCCCTTAGCAAAAAAACCCCTGCACCTTACTGTGCTATTATATATCC 1005
||||: |||||: |||||: |||||: |||||: |||
Db 1401 ATGLeuIleGlyLeuLysGlnPheAsnAspTyrHisMet-----SerylSerLys 1417
QY 1006 TTAGTTCCAAGTTAATATATCTTATTTCTGGATATTTGCTTTATACCAAGAGCTTATC 1065
||| ||: ||: |||||: ||| ||| |||
Db 1418 SerGlnThrLysGlnMetSerAsnPheIleAspSerylTyrPheValLysProIle--- 1436
QY 1066 AGCCAGTTCACGAACACCACTATPACGCAACAACATACCAACACATACCAATATA 1125
1437 --AsnMetAspCysAlaHisTyr-----Ile 1444
QY 1126 CGTTTAAATATCATC-----AGTAACTCAGACATGATTTAGAGCTTGATGGCAA 1179
::: |||||: |||||: |||||: |||||: |||||: |||
Db 1445 LysHisGluLeuValLeuTyrAsnAspLeuIleThrHisPheIleSerLeuAsnLeuTyr 1464
QY 1180 ATTAGCATTTCATACATCCATATCTCATCTTCA---TACCATATACACTACTACCACTT 1236
||| ||| |||||: ||| ||| ||| |||
Db 1465 ILeSerAspGlnHisGlyLeuLysSerLeuSerIleAsnIleLeuLysIle 1484
QY 1237 TTGTGATCATCTTAAGACCAATGCAGATTAACCCCTATATATTA 1284
||| |||||: |||||: |||||: |||||: |||||: |||
Db 1485 PheIleThrSerAsnGlnSerGlnIleSerLeuAlaProIleGlyIle 1500

RESULT 13
A53611
Interleukin-8 receptor type B - human
C:Species: Homo sapiens (man)
C:Date: 07-Oct-1994 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
R:Ahuja, S.K.; Shetty, A.; Tiffany, H.L.; Murphy, P.M.
J. Biol. Chem. 269, 26381-26389, 1994
A:Title: Comparison of the genomic organization and promoter function for human Interleu

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-360 <RES>
A:Cross-references: EMBL:U11872; NID:9511801; PIDN:AAB60656.1; PID:9511803
A:Accession: J18712

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-15 <RE2>
A:Cross-references: EMBL:U11872; NID:9511808; PIDN:AAA64380.1; PID:9511809; EMBL:U11877; NID:9511816; PID:9511817; EMBL:U11877; NID:9511818; PID:9511819; EMBL:U11878; J.Biol.Chem. 265, 11065-11072, 1994

A>Title: Structure, genomic organization, and expression of the human interleukin-8 r
A:Reference number: A53611; MUID:94209273; PMID:7512557
A:Accession: A53611

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 6-360 <SPR>

A:Cross-references: GB:M99412; GB:L19593
Murphy, P.M.; Tiffany, H.L.
Science 253, 1280-1283, 1991

A>Title: Cloning of complementary DNA encoding a functional human Interleukin-8 recep
A:Reference number: A39446; MUID:91368200; PMID:1891716
A:Accession: A39446

A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 6-360 <MUR>
A:Cross-references: GB:M73969
C:Comment: This receptor, unlike IL8RA, binds several peptides besides interleukin-8,
C:Genetics:
A:Gene: GDB:IL8RB; IL8RA
A:Cross-references: GDB:127868; OMIM:146928
A:Map position: 2q35-2q35
C:Superfamily: vertebrate rhodopsin
C:Keywords: G protein-coupled receptor; transmembrane protein

Alignment Scores:

Pred. No.:	1-77	Length:	360
Score:	85.50	Matches:	58
Percent Similarity:	37.84%	Conservative:	40
Best Local Similarity:	22.39%	Mismatches:	87
Query Match:	3.70%	Indels:	75
DB:	2	Gaps:	10

US-09-030-244B-2 (1-1331) x A53611 (1-360)

QY 1187 AGTCGATTTCCCATTAACCTCAATAATCATGTCTCGACTTACTGATG-----ATA 1134
|||||
Db 76 SerAlvalGLyargSerValThraspValItyrLeuleuSlnLeuAlalaLeuAlaspleu 95
|||
QY 1133 TTAAAGCATATATGGTGATGTGGTATGTGTGTTGCTATAGTGCTGTTGCTGG 1074
|||
Db 96 LeuhenAlaleuThrLeuProileTrpAlaAlaserLyValasnGlyrrPlepegly 115
|||
QY 1073 AACGCGCTGATTAAGGCTCTTGGTATTAAGAACAATTCAGAAATTAAGTATTAACTTT 1014
:::|
Db 116 ThrpheLeu-----CysLysValValserLeuLeuLysluValasnPh 130
:::|
QY 1013 GAACATAAGGGGATATTAATAGCACACAGTAAAGGTGAAGGCTTTTTTTCTTAAGGAG 954
:::|
Db 131 -----TyrSerGly----- 133

QY 953 ACAGAGTTTATTATTACTCAATCAGTCTGCCA-AAAGATTAATTAAATCGCTTT 895
:::|
Db 134 -----lleLeuLeuLeuAlacyslleserValasphAgtyrLeuAlallevAlhis 150
:::|
QY 894 TCPTCCACACTTTTGAACAACACACTTTCGTCATGTTGCATGACAGGGGACTATCTGACAGAAAT 835
:::|
Db 151 AlaThrArg-----ThrLeuThrInLynsArgTyrrLeuValLysphe 164
:::|
QY 834 ATTTGT-----CTGTGACATCAGTACTTAAGTCTTTAAAAAAA 796
|||||
Db 165 IllecylseuSerilerIpolyLeuSerleuLeuLeuAlalaLeuproValleuLeuPhearg 184

489 TFC

;Molecule type: protein
;Accession: B32524

```
;Accession: B32524
;Molecule type: protein
;Residues: 1-19 <OP1>
;Note: 6-Leu was also found
```

R:Oppenheim, F.G.; Offner, G.D.; Troxler, R.F.
 J. Biol. Chem. 257, 9271-9282, 1982
 A:Title: Phosphoproteins in the parotid saliva from the subhuman primate Macaca fascicularis
 of a proline-rich phosphopeptide.
 R:Reference number: A03289; MUID:82265555; PMID:7107568
 A:Accession: A03289
 A:Molecule type: protein
 A:Residues: 20-61 <OP2>
 C:Superfamily: statherin precursor; statherin/histatin signal sequence homology
 C:Keywords: phosphoprotein; saliva
 F:1-25/Domain: statherin/histatin signal sequence homology <SHH>
 F:1-19/Domain: signal sequence #status experimental <SIG>
 F:20-61/Product: statherin #status experimental <MAT>
 F:21,22/Binding site: phosphate (Ser) (covalent) #status experimental

Alignment Scores:

Pred. No.:	1.64	Length:	61
Score:	85.00	Matches:	15
Percent Similarity:	68.18%	Conservative:	0
Best Local Similarity:	68.18%	Mismatches:	7
Query Match:	3.66%	Indels:	0
DB:	1	Gaps:	0

US-09-830-244b-2 (1-1331) x SBMOP1 (1-61)

OY	1059	CCTTATCAGCCAGTTCAGAACACACATATACGACACACCAATACCAACACACATACCAA	1118
Db	40	ProTyrGlnProPheAlaProGlnProLeuTyrProGlnProTyrGlnPro	59
OY	1119	CAATAT	1124
Db	60	GlnTyr	61

Search completed: June 24, 2003, 19:25:33
 Job time : 64 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 24, 2003, 19:15:17 ; Search time 21 Seconds

(without alignments)
5257.618 Million cell updates/sec

Title: US-09-830-244b-2

Perfect score: 2322
Sequence: 1 ctatgtttttagaatcaag.....gccttttcaatgactctg 1331

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL-frame+.n2p.model -DEV-xlp
-Q=/cgn2./USPTO.spool/US09830244.rnat.24062003.130001.10030/app.query.fasta.1.1479
-DB=SwissProt.40 -OPMT-fastan -SUPRT-1 -BND-1 -MATRIX-biosum62 -TRANS-human40.cdi
-LOOEXT-0 -UNITS-bits -SPART-1 -EPR-1 -MATRIX-biosum62 -TRANS-human40.cdi
-LIST-45 -DOCATIGN-200 -THR_SCORE-pct -THR_MAX-100 -THR_MIN-0 -ALIGN-15
-MODE-LOCAL -OUTFMT-pto -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000
-USER-US09830244.ecn.1.1.30.grnat.24062003.130001.10030 -NCPU-6 -ICPU-3
-NO_MAP -LARGEQUERY -NEG_SCORES-0 -WAIT -DSPBLOCK-100 -LONGLOG
-DEV_TIMEOUT-120 -WARN_TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-6
-FCGAPEXT-7 -YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEX-7

Database: SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	5.9	62	1	STAT_HUMAN
2	91.5	4.0	1138	1	P02808 homo sapien
3	90	3.9	853	1	P3037 saccharomyc
4	89	3.8	467	1	Q95V4 homo sapien
5	88.5	3.8	353	1	IL8B_PHYPO
6	85.5	3.7	360	1	IL8B_MACMU
7	85.5	3.7	1070	1	EXSC_BUCAT
8	85	3.7	61	1	STAT_HUMAN
9	85	3.7	320	1	STAT_HUMAN
10	85	3.7	567	1	NO2M_HAWMI
11	84	3.6	378	1	EDG3_HUMAN
12	84	3.6	447	1	TBB1_CYPAP
13	83.5	3.6	353	1	IL8B_PANTR
14	83	3.6	442	1	TBB1_PANTR
15	83	3.6	1037	1	N120_YEAST
16	82.5	3.6	362	1	ERD1_YEAST
17	82.5	3.6	476	1	ATXA_BACAN
18	81	3.5	835	1	Y4722_MYCGE

19	80	3.4	445	1	TBB1_COLGR	P22013 colleotric
20	80	3.5	445	1	Y148_MYCPN	P75584 mycoplasma
21	79.5	3.4	482	1	RNF9_HUMAN	Q9udv6 homo sapien
22	79.5	3.4	572	1	YB30_YEAST	P38125 saccharomyc
23	79.5	3.4	635	1	ETFL_YABAM	Q9gdb7 yaba monkey
24	79.5	3.4	1329	1	FTSK_ECOLI	P46889 escherichia
25	79.5	3.4	1612	1	RRPO_PPMYS	P89657 pepper mild
26	79.5	3.4	1612	1	RRPO_PPMYS	P29098 pepper mild
27	79	3.4	359	1	IL8B_RAT	P35407 rattus norv
28	79	3.4	447	1	TBB2_PEA	P29501 pistum saliv
29	78.5	3.4	353	1	IL8B_GORGO	Q28422 gorilla gor
30	78	3.4	42	1	STAT_MACAR	P14709 mecate arct
31	77.5	3.4	350	1	IL8A_PANTR	P55920 pan troglod
32	77.5	3.4	483	1	PRR_BUCAT	P57386 buchiera ap
33	77.5	3.4	1237	1	POL4_DROME	P10394 drosophila
34	77	3.3	262	1	MURI_BUCAT	P57619 buchiera ap
35	77	3.3	417	1	IF_HUMAN	P27352 homo sapien
36	77	3.3	443	1	TBB1_THAME	O91K18 thalassiosl
37	77	3.3	451	1	TBB2_SOLITU	P46263 solanum tub
38	77	3.3	452	1	GP11_SCHPO	P46264 solanum tub
39	77	3.3	653	1	YBTF6_YEAST	O14357 schizosacch
40	77	3.3	946	1	YBTF6_YEAST	P38250 saccharomyc
41	76.5	3.3	350	1	IL8A_HUMAN	P25024 homo sapien
42	76	3.3	237	1	YFV4_METTE	P29575 methanobact
43	76	3.3	359	1	IL8B_RAT	P35407 rattus norv
44	76	3.3	760	1	ALK1_YEAST	P43633 saccharomyc
45	75.5	3.3	439	1	SYE2_HELPY	O25360 helicobacte

ALIGNMENTS

RESULT 1	STAT_HUMAN	STANDARD	PRT	62 AA
ID	STAT_HUMAN			
AC	P02808			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-JUL-1989 (Rel. 11, Last sequence update)			
DT	01-AUG-1991 (Rel. 19, Last annotation update)			
DE	Statherin precursor.			
GN	STAT_H.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=8074310; PubMed=3502720;			
RA	Sabatini L., Carlock L., Johnson G., Azen E.;			
RT	"CDNA cloning and chromosomal localization (4q11-13) of a gene for			
RT	statherin, a regulator of calcium in saliva."			
RL	Am. J. Hum. Genet. 41:1048-1060(1987).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=88106506; PubMed=3426601;			
RA	Dickinson D.P., Ridall A.L., Levine M.J.;			
RT	"Human submandibular gland statherin and basic histidine-rich peptide			
RT	are encoded by highly abundant mRNA's derived from a common ancestral			
RT	sequence."			
RL	Biochem. Biophys. Res. Commun. 149:784-790(1987).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=90323623; PubMed=2373369;			
RA	Sabatini L.M., He T., Azen E.A.;			
RT	"Structure and sequence determination of the gene encoding human			
RT	salivary statherin."			
RL	Gene 89:245-251(1990).			
RN	[4]			
RP	SEQUENCE OF 20-62.			
RP	MEDLINE=77118656; PubMed=838735;			
RA	Schlesinger D.H., Hay D.I.;			
RT	"Complete covalent structure of statherin, a tyrosine-rich acidic			
RT	peptide which inhibits calcium phosphate precipitation from human			
RT	parotid saliva."			


```

US-09-830-244b-2 (1-1331) x DPS2_YEAST (1-1138)
OY 927 AGTCTCCCAAAAGTACTTAAATGCGTTTCTCCAC-----TTTGAACCA 874
DB 751 ThrleuLysProSerTyrValSerSerAlaPheAlaSerLysArgTyrPheSerLysVal 770
OY 873 CATTTGGTACTGTTGACAGGGGACTATCTGATC-----AGAAATTTTGTCTG 826
DB 771 LeuIleLysLeuThrTyrGlyLysTyrAlaLeuGlySerAsnAsnAlaAsnIleLeuVal 790
OY 825 TTGACATGAGTACTTAAGTCTTTAAACAAAAAAGAGCGCTTCAGATTAAGAAATTAGG 766
DB 791 GlnAspArgAspPheGlyIleIleIleGlnLysIleSerAlaIleValLysLeuGly 810
OY 765 GCTCTAATAGATGTTATATGATCATATGATGCTCTTTTGTAGACAGAGTCTGTCTTA 706
DB 811 MetArgIle-----IleTyrAsnGlyLysSerPro----- 820
OY 705 AAAAACAACAAAGAAATGAGAGTTCTACTAATGCTCTGCTCCAGAGGGGATTC 646
DB 821 -----GlnSerLysLysPheArgSerLeuLeuLysThrLeuSerIleArgGlnGly--- 837
OY 645 AATGAGATGCTAAGCAAGCAAAACACCTGCAGGGGAGAGCAATCTTGATTTTCAGATA 586
DB 838 -----LysLysPheAspSerThrAlaSerAlaLysGlnIleGlnProPhe---Ile 853
OY 585 AAGTACCAAAATAATATAGTATCTCTTTCGACAGATGCGCTTCTCTGTTAAACT 526
DB 854 LysPheHisSer-----LeuAspLeuSerGlnCysArgTyrAspLysAspPhe---LysThr 870
OY 525 TTGAGTATTCACAAAGCGCTATGTTGAGAGTGCAGCAATCCAGTACAGTTACTAGTA 466
DB 871 PheAsnGln-----PhePheTyrArgLysLeuLysProGlySerArgLeuProGlu 887
OY 465 ACAGTTATTACTATGCTATGCTATTCACATTAACGCGATTACCAATTAAGCATTAGTG 406
DB 888 SerAsnAsnLysGlnIleLeuPheSerPro----- 897
OY 405 CACATCAAAAGATCTGAAATGCTAATGGAGCATTTCCCAAAAGAAATCCCAAT 346
DB 898 -----AlaAspSerArgCysThrValPheProThrIleGlnGlnSerLysGlu 913
OY 345 GCGTCTATTAAGGAGCAATATGCAATGGCAAGCAAGATGAATTTCTGTCAAAAGAA 286
DB 914 IleTyrValLysGlyArg-----LysPheSerIleLysLys 925
OY 285 AGTACTCTCAAGCAATCAGAAAAACTTCCCAAT 247
DB 926 LeuAlaAsnAsnTyrAsnProGluThrPheAsnAspAsn 938

RESULT 3
DIA3_HUMAN STANDARD; PRT: 853 AA.
ID DIA3_HUMAN
AC 09NSV4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Diaphanous protein homolog 3 (Diaphanous-related formin 3) (DRF3)
DE (Fragment).
GN DIA3 OR DIA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Testis;
RA Blum H., Buersachs S., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RM [2]
RP SEQUENCE OF 468-853 FROM N.A. (ISOFORMS 1 AND 2).
RA Smith M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

```

```

CC -1- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN
CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE
CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,
CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
CC SERUM RESPONSE FACTOR. DRF PROTEINS COUPLE RHO AND SRC TYROSINE
CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS (BY
CC SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
CC RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GBD-DAD INTERACTION
CC (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DRF AUTOREGULATORY DOMAIN (DAD).
CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY, DIAPHANOUS
CC SUBFAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: AL137718; CAB70890.1; ALT_INTR.
DR EMBL: AL354829; CAC17664.1;
DR EMBL: AL354829; CAC17665.1;
DR Genew; HGNC:115480; DIA3.
DR Interpro; IPR003104; FH2.
DR Pfam; PF02181; FH2; 1.
DR SMART; SM00498; FH2; 1.
DR SMART; SM00498; FH2; 1.
KW Coiled coil; Repeat; Alternative splicing.
FT NON_TER 1
FT DOMAIN 304 374 FH1 (PRO-RICH).
FT DOMAIN 379 818 FH2.
FT DOMAIN 801 815 DAD.
FT DOMAIN 137 167 COILED COIL (POTENTIAL).
FT DOMAIN 241 299 COILED COIL (POTENTIAL).
FT DOMAIN 650 799 COILED COIL (POTENTIAL).
FT DOMAIN 819 822 ARG/LYS-RICH (BASIC).
FT DOMAIN 656 697 VSVETLEKNDKRWGROLOOLEKLETFEPPEDEHDKFYKI
FT F -> GICLPEKHPALISAKRLKIPICMTFPLSHVF
FT FT IPRISF (IN ISOFORM 2).
FT FT MISSING (IN ISOFORM 2).
SQ SEQUENCE 853 AA; 98610 MW; B7FA9C745AE18CD9 CRC64;

```

```

Alignment Scores:
Pred. No.: 0.427 Length: 853
Score: 90.00 Matches: 46
Percent Similarity: 40.708 Conservative: 35
Best Local Similarity: 23.128 Mismatches: 63
Query Match: 3.898 Indels: 55
DB: 1 Gaps: 9

```

```

US-09-830-244b-2 (1-1331) x DIA3_HUMAN (1-853)
OY 717 GTCCTGCTTAAACAAACAAAGAAATGAGAGTTCTACTA-----ACT 670
DB 409 IleLysValAsnGlnAlaAsnLysTyrGlnAsnValAspLeuLeuCysLysLeuGlnAsnThr 428
OY 669 GTCCTGCTGCCAGAGGGGATTCAATAGAGTCTTAAGCAAGAAACACCTGCAGGG 610
DB 429 PheCysCysGlnGlnLys-----GlnArgArgGlnGlnGlnAspIleGlnGlu 444
OY 609 AAGCAGACTCTTGATTCAGATAAAGTACCAAAATAATATAGCATCTCTTCTGCAGCA 550
DB 445 LysLysSerIleLysLysLysIleLysGlnLeuLysPheLeu----- 458

```

OY	549	CATGCCCTTCTTCGTGTTAAACACTTTCAGTCATTCACAAAGGCTATTGTAGAGTGCACG	450
Db	458	-----	458
OY	489	AAATCCAGTGTAGATTTTACTGTACAGAGTTATTTAACTATGCTGTATTCACCAATACG	430
Db	459	AspSerLysIleValIaGlnAsnLeuSerIlePheLeuSerSerPheArgValIProTyrGlu	478
OY	429	GCATTACCAATAAATGCGATT-----GTGCACATCAAAAGATCT-----GAA	388
Db	479	GluIleArgMetMetIleLeuGluValAlaSpGIunThrArgLeuAlaGluSerMetIleGln	498
OY	387	AATCCTAATGGAGACATTTCCCAAAAGAAAGAAAG-----AATCCATGCGCTCTTAAAGG	331
Db	499	Asn-LeuIleLysHisLeuProAspGlnGluInLeuAsnSerLeuSerGlnPheLysSe	518
OY	330	AGAATATAGCAATGGCAAGCAAGATGAATAATTTCTGTCAAAAGAAAGTACTTCTCAACG	271
Db	518	rgLTuTyrSerAsnLeuGlySgluProGlnInPheVal-----Va	531
OY	270	AATCAGAGAAACAGTGTCCCAATTTAGCGCTGGGAGACCATTAACCATATCTCAGAT	211
Db	531	ValMetSerAsnValLysArgLeuArgPro-----ArgLeuSerAlaI	546
OY	210	ACTGATTAATTG---TTTAGCCTTAGGGGCGCTCAATATTTCCCGCATCGCTA	159
Db	546	eleuPheLysLeuGlnPheGluGluGlnValAsnAsnIleLysProAspIleMet	564
RESULT 4			
ID	TBBL_PHYPO	STANDARD;	PRT; 467 AA.
AC	P07436;		
DT	01-APR-1988 (Rel. 07, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DE	Tubulin beta-1 chain.		
DE	BETA AND B2BP.		
OS	Physarum polycephalum (Slime mold).		
OC	Eukaryota; Mycetozoa; Myxogastria; Myxogastromycetidae; Physarida;		
OC	Physarum.		
ON	NCBI_TaxId=5791;		
RX	SEQUENCE FROM N.A.		
RC	STRAIN-LU352;		
RA	Paul E.C.A., Buchsacher G.L. Jr., Cunningham D.B., Dove W.F.,		
RA	Burland T.G.,		
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE OF 1-445 FROM N.A.		
RC	STRAIN-LU352;		
RX	MEDLINE=92211323; PubMed=1556551;		
RA	Paul E.C.A., Buchsacher G.L. Jr., Cunningham D.B., Dove W.F.,		
RA	Burland T.G.,		
RT	"Preferential expression of one beta-tubulin gene during flagellate		
RT	development in Physarum.";		
RL	J. Gen. Microbiol. 138:229-238(1992).		
RN	[3]		
RP	SEQUENCE OF 4-207 FROM N.A.		
RC	STRAIN-CL;		
RX	MEDLINE=88271316; PubMed=3391166;		
RA	Wetenskiold A.K., Poetsch B., Haugli F.,		
RA	"Cloning and expression of a beta tubulin gene of Physarum		
RT	polycephalum.";		
RT	Eur. J. Biochem. 174:491-495(1988).		
RN	[4]		
RP	PARTIAL SEQUENCE OF 1-217: 234-262 AND 277-286.		
RX	MEDLINE=87080317; PubMed=3539596;		
RA	Slingshoter-Morra W., Clayton L., Dawson P., Gull K., Little M.,		
RA	"Amino-acid sequence data of beta-tubulin from Physarum polycephalum		
RT	myxamebebe.";		
RT	Eur. J. Biochem. 161:669-679(1986).		
RL	-1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT		
CC	BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA		

```

CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
CC -1 SUBUNIT: DIMER OF ALPHA AND BETA CHAINS.
CC -1 SUBCELLULAR LOCATION: MITOSIS IN THE SLIME MOLD PLASMODIUM DIFFERS
CC FROM THE PROCESS IN MANY EUKARYOTES. THE TUBULIN CHAINS MUST BE
CC TRANSPORTED TO THE NUCLEI FOR INTRANUCLEAR ASSEMBLY OF THE
CC SPINDLE.
CC -1 DEVELOPMENTAL STAGE: BETA IS PREFERENTIALLY EXPRESSED IN
CC FLAGELLATE AND BETA IN AMOEBA.
CC -1 SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M58521; AAA29974.1;
DR EMBL: X12371; CA30932.1;
DR PIR: A44849; A44848.
DR PIR: A25342; A25342.
DR PIR: S02532; S02532.
DR InterPro: IPR002453; Beta_tubulin.
DR InterPro: IPR000217; Tubulin.
DR InterPro: IPR003008; Tubulin_FtsZ.
DR Pfam: PF00091; tubulin_1.
DR PRINTS: PR0161; TUBULIN.
DR PROSITE: PS00227; TUBULIN.
DR PROSITE: PS00228; TUBULIN_B_AUTOREG; 1.
KW Microtubules; GTP-binding; Multigene family; Nuclear protein.
FT NP_BIND 140 146 GTP (POTENTIAL).
FT VARIANT 39 39 E -> D (IN BETB).
FT VARIANT 283 283 S -> A (IN BETB).
FT CONFLICT 165 165 C -> D (IN REF. 4).
FT CONFLICT 196 196 A -> T (IN REF. 4).
FT CONFLICT 238 238 C -> S (IN REF. 4).
SQ SEQUENCE 467 AA; 52134 MW; BA2C330A65FDC964 CRC64;

Alignment Scores:
Pred. No.: 0 505
Score: 89 00 Length: 467
Percent Similarity: 39.81% Matches: 44
Best Local Similarity: 21.36% Conservative: 36
Query Match: 3.83% Mismatches: 78
DB: Indels: 48
Gaps: 8

US-09-830-244B-2 (1-1331) x TBB1_PHYPO (1-467)

OY 386 TTTTCAGATGTTTGTGACACTAATGACATATTGTAATGCCGTTATGTCGAATAC 445
DB :::: ||| ::::::::::::::: ::::::::::: :::::
OY 159 TYPTRASPTARGMETWELCYSTRPHESEIRVALValProserProlySvalSerAsPThr 178
OY 446 AGATAGTAAATAAACGTGTACACTAATCTACAC----- 481
DB ::::: ||| ::::::::::::::: ||| ::::::::::::::: |||
OY 179 ValValGlpProTyrAsnAlaThrIleuSerValHisGlnIleuValGluAsnAlaAspGlu 158
OY 482 -----TTGCATTTGCTGCACCTCTACCAATAGCCTTTTGATGAGTGAAGAGTTTA 532
DB ::::: ||| ||||| ||||| ||| |||||
OY 199 ValMetCysIleAspAsnGluAlaIleuValYrAspIleSerPheArgThrIleuSleuThr 218
OY 533 ACAGAGAAAGAGAGCATGTCGACAGAAAGATACTAATATTTTGTGCTACTTATCTGA 592
DB ::::: ||| ::::::::::: ||| :::::::::::
OY 219 ThrProThrTyrGlyAspLeuAsnHisIleuValSerAlaVal-----MetSer 234
OY 593 AATCCAAAGTCTGTC-----TTCCCTCGCAGGTTGTTTCTCTCTTACGATCC----- 640
DB ::::: ||||| ||||| :::: ||| |||||
OY 235 GlyIleThrCysCysLeuArgPheProGlyGlnIleuAsnSerAspLeuArgIleuAla 254
OY 641 ----TATTGAATGCCCTCGGAGACACAGACAGACTTGTAGACTCTCCATTTCTTTGTT 697
DB ::::: ||| ||| ||||| ||||| ||||| ||||| |||||
OY 255 ValAlaIleuIleProPhe-----ProArgIleuHisIlePheIleu 267

```

QY 698 TGTGTTTAAAGACAGACTGTCTCAAAAAAGCATTTATCATTA----- 748
 Db 268 ValGlyPheLeuAlaProLeuHisArgGlySerValGlyTyrArgSerLeuThrValPro 287
 QY 749 -----TAACTTTTATTAGAGCCCTTAAT----- 772
 Db 288 GLeuLeuThrGlnGlnMetPheAspAlaLysAsnMetValAlaSerAspProArgHis 307
 QY 773 -----TTCTATCTGAAAGCAGCTGTTTTTTTAAACAGTAACTAGTACGATGCAAC 826
 Db 308 GlyArgTyrLeuThrAlaSerAlaMetPheArgGlyArgMetSerThrLysGluValAlaSp 327
 QY 827 AGACAAATTTTCTGATCAGATAGTCCCTGTCAACAGTCAATGCGTTTCATAAG 886
 Db 328 GlnGlnMetLeuAsnValGln-----AsnLysAsnSerSerTyrPheValGlu 343
 QY 887 TGGGAGAAAGACAGCATT 904
 Db 344 TrpIleProAsnAsnIle 349

RESULT 5
 IL8B_MACMU STANDARD: PRT: 353 AA.
 AC Q28519;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE High affinity Interleukin-8 receptor B (IL-8R B) (CXCR-2) (Fragment).
 GN IL8RB OR CXCR2.
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 NX NCBI_TaxID:9544;
 [1]
 RE SEQUENCE FROM N.A.
 RA MEDLINE:96175151; PubMed:9110929;
 RA Alvarez V., Coco E., Setten F., Gonzalez S., Gonzalez-Roces S.,
 RA Lopez-Larrea C.;
 RT Characterization of interleukin-8 receptors in non-human primates.;
 RL Immunogenetics 43:261-267(1996).
 CC - FUNCTION: RECEPTOR TO INTERLEUKIN-8, WHICH IS A POWERFUL
 CC NEUTROPHILS CHEMOTACTIC FACTOR. BINDING OF IL-8 TO THE RECEPTOR
 CC CAUSES ACTIVATION OF NEUTROPHILS. THIS RESPONSE IS MEDIATED VIA A
 CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
 CC MESSENGER SYSTEM. THIS RECEPTOR BINDS TO IL-8 WITH A HIGH AFFINITY
 CC AND TO GROMS/A AND NAP-2 ALSO WITH A HIGH AFFINITY.
 CC - SUBCELLULAR LOCATION: Integral membrane protein.
 CC - SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X91116; CA662565.1;
 DR HSP: P34996; JDD;
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR InterPro: IPR000832; GPCR_secretin.
 DR Pfam: PF00001; 7tm_1; 1.
 DR PROSITE: PS00237; G-PROTEIN_RECEP_P1_1; 1.
 DR PROSITE: PS50262; G-PROTEIN_RECEP_P1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW Chemotaxis.
 FT NON_TER 1
 FT DOMAIN 1
 FT TRANSMEM 45
 FT TRANSMEM 46
 FT TRANSMEM 72
 FT TRANSMEM 73
 FT TRANSMEM 81
 FT TRANSMEM 82
 FT TRANSMEM 102
 FT TRANSMEM 103
 FT TRANSMEM 117
 FT TRANSMEM 127
 FT TRANSMEM 130
 FT TRANSMEM 131
 FT TRANSMEM 132
 FT TRANSMEM 133
 FT TRANSMEM 134
 FT TRANSMEM 135
 FT TRANSMEM 136
 FT TRANSMEM 137
 FT TRANSMEM 138
 FT TRANSMEM 139
 FT TRANSMEM 140
 FT TRANSMEM 141
 FT TRANSMEM 142
 FT TRANSMEM 143
 FT TRANSMEM 144
 FT TRANSMEM 145
 FT TRANSMEM 146
 FT TRANSMEM 147
 FT TRANSMEM 148
 FT TRANSMEM 149
 FT TRANSMEM 150
 FT TRANSMEM 151
 FT TRANSMEM 152
 FT TRANSMEM 153
 FT TRANSMEM 154
 FT TRANSMEM 155
 FT TRANSMEM 156
 FT TRANSMEM 157
 FT TRANSMEM 158
 FT TRANSMEM 159
 FT TRANSMEM 160
 FT TRANSMEM 161
 FT TRANSMEM 162
 FT TRANSMEM 163
 FT TRANSMEM 164
 FT TRANSMEM 165
 FT TRANSMEM 166
 FT TRANSMEM 167
 FT TRANSMEM 168
 FT TRANSMEM 169
 FT TRANSMEM 170
 FT TRANSMEM 171
 FT TRANSMEM 172
 FT TRANSMEM 173
 FT TRANSMEM 174
 FT TRANSMEM 175
 FT TRANSMEM 176
 FT TRANSMEM 177
 FT TRANSMEM 178
 FT TRANSMEM 179
 FT TRANSMEM 180
 FT TRANSMEM 181
 FT TRANSMEM 182
 FT TRANSMEM 183
 FT TRANSMEM 184
 FT TRANSMEM 185
 FT TRANSMEM 186
 FT TRANSMEM 187
 FT TRANSMEM 188
 FT TRANSMEM 189
 FT TRANSMEM 190
 FT TRANSMEM 191
 FT TRANSMEM 192
 FT TRANSMEM 193
 FT TRANSMEM 194
 FT TRANSMEM 195
 FT TRANSMEM 196
 FT TRANSMEM 197
 FT TRANSMEM 198
 FT TRANSMEM 199
 FT TRANSMEM 200
 FT TRANSMEM 201
 FT TRANSMEM 202
 FT TRANSMEM 203
 FT TRANSMEM 204
 FT TRANSMEM 205
 FT TRANSMEM 206
 FT TRANSMEM 207
 FT TRANSMEM 208
 FT TRANSMEM 209
 FT TRANSMEM 210
 FT TRANSMEM 211
 FT TRANSMEM 212
 FT TRANSMEM 213
 FT TRANSMEM 214
 FT TRANSMEM 215
 FT TRANSMEM 216
 FT TRANSMEM 217
 FT TRANSMEM 218
 FT TRANSMEM 219
 FT TRANSMEM 220
 FT TRANSMEM 221
 FT TRANSMEM 222
 FT TRANSMEM 223
 FT TRANSMEM 224
 FT TRANSMEM 225
 FT TRANSMEM 226
 FT TRANSMEM 227
 FT TRANSMEM 228
 FT TRANSMEM 229
 FT TRANSMEM 230
 FT TRANSMEM 231
 FT TRANSMEM 232
 FT TRANSMEM 233
 FT TRANSMEM 234
 FT TRANSMEM 235
 FT TRANSMEM 236
 FT TRANSMEM 237
 FT TRANSMEM 238
 FT TRANSMEM 239
 FT TRANSMEM 240
 FT TRANSMEM 241
 FT TRANSMEM 242
 FT TRANSMEM 243
 FT TRANSMEM 244
 FT TRANSMEM 245
 FT TRANSMEM 246
 FT TRANSMEM 247
 FT TRANSMEM 248
 FT TRANSMEM 249
 FT TRANSMEM 250
 FT TRANSMEM 251
 FT TRANSMEM 252
 FT TRANSMEM 253
 FT TRANSMEM 254
 FT TRANSMEM 255
 FT TRANSMEM 256
 FT TRANSMEM 257
 FT TRANSMEM 258
 FT TRANSMEM 259
 FT TRANSMEM 260
 FT TRANSMEM 261
 FT TRANSMEM 262
 FT TRANSMEM 263
 FT TRANSMEM 264
 FT TRANSMEM 265
 FT TRANSMEM 266
 FT TRANSMEM 267
 FT TRANSMEM 268
 FT TRANSMEM 269
 FT TRANSMEM 270
 FT TRANSMEM 271
 FT TRANSMEM 272
 FT TRANSMEM 273
 FT TRANSMEM 274
 FT TRANSMEM 275
 FT TRANSMEM 276
 FT TRANSMEM 277
 FT TRANSMEM 278
 FT TRANSMEM 279
 FT TRANSMEM 280
 FT TRANSMEM 281
 FT TRANSMEM 282
 FT TRANSMEM 283
 FT TRANSMEM 284
 FT TRANSMEM 285
 FT TRANSMEM 286
 FT TRANSMEM 287
 FT TRANSMEM 288
 FT TRANSMEM 289
 FT TRANSMEM 290
 FT TRANSMEM 291
 FT TRANSMEM 292
 FT TRANSMEM 293
 FT TRANSMEM 294
 FT TRANSMEM 295
 FT TRANSMEM 296
 FT TRANSMEM 297
 FT TRANSMEM 298
 FT TRANSMEM 299
 FT TRANSMEM 300
 FT TRANSMEM 301
 FT TRANSMEM 302
 FT TRANSMEM 303
 FT TRANSMEM 304
 FT TRANSMEM 305
 FT TRANSMEM 306
 FT TRANSMEM 307
 FT TRANSMEM 308
 FT TRANSMEM 309
 FT TRANSMEM 310
 FT TRANSMEM 311
 FT TRANSMEM 312
 FT TRANSMEM 313
 FT TRANSMEM 314
 FT TRANSMEM 315
 FT TRANSMEM 316
 FT TRANSMEM 317
 FT TRANSMEM 318
 FT TRANSMEM 319
 FT TRANSMEM 320
 FT TRANSMEM 321
 FT TRANSMEM 322
 FT TRANSMEM 323
 FT TRANSMEM 324
 FT TRANSMEM 325
 FT TRANSMEM 326
 FT TRANSMEM 327
 FT TRANSMEM 328
 FT TRANSMEM 329
 FT TRANSMEM 330
 FT TRANSMEM 331
 FT TRANSMEM 332
 FT TRANSMEM 333
 FT TRANSMEM 334
 FT TRANSMEM 335
 FT TRANSMEM 336
 FT TRANSMEM 337
 FT TRANSMEM 338
 FT TRANSMEM 339
 FT TRANSMEM 340
 FT TRANSMEM 341
 FT TRANSMEM 342
 FT TRANSMEM 343
 FT TRANSMEM 344
 FT TRANSMEM 345
 FT TRANSMEM 346
 FT TRANSMEM 347
 FT TRANSMEM 348
 FT TRANSMEM 349
 FT TRANSMEM 350
 FT TRANSMEM 351
 FT TRANSMEM 352
 FT TRANSMEM 353
 FT TRANSMEM 354
 FT TRANSMEM 355
 FT TRANSMEM 356
 FT TRANSMEM 357
 FT TRANSMEM 358
 FT TRANSMEM 359
 FT TRANSMEM 360
 FT TRANSMEM 361
 FT TRANSMEM 362
 FT TRANSMEM 363
 FT TRANSMEM 364
 FT TRANSMEM 365
 FT TRANSMEM 366
 FT TRANSMEM 367
 FT TRANSMEM 368
 FT TRANSMEM 369
 FT TRANSMEM 370
 FT TRANSMEM 371
 FT TRANSMEM 372
 FT TRANSMEM 373
 FT TRANSMEM 374
 FT TRANSMEM 375
 FT TRANSMEM 376
 FT TRANSMEM 377
 FT TRANSMEM 378
 FT TRANSMEM 379
 FT TRANSMEM 380
 FT TRANSMEM 381
 FT TRANSMEM 382
 FT TRANSMEM 383
 FT TRANSMEM 384
 FT TRANSMEM 385
 FT TRANSMEM 386
 FT TRANSMEM 387
 FT TRANSMEM 388
 FT TRANSMEM 389
 FT TRANSMEM 390
 FT TRANSMEM 391
 FT TRANSMEM 392
 FT TRANSMEM 393
 FT TRANSMEM 394
 FT TRANSMEM 395
 FT TRANSMEM 396
 FT TRANSMEM 397
 FT TRANSMEM 398
 FT TRANSMEM 399
 FT TRANSMEM 400
 FT TRANSMEM 401
 FT TRANSMEM 402
 FT TRANSMEM 403
 FT TRANSMEM 404
 FT TRANSMEM 405
 FT TRANSMEM 406
 FT TRANSMEM 407
 FT TRANSMEM 408
 FT TRANSMEM 409
 FT TRANSMEM 410
 FT TRANSMEM 411
 FT TRANSMEM 412
 FT TRANSMEM 413
 FT TRANSMEM 414
 FT TRANSMEM 415
 FT TRANSMEM 416
 FT TRANSMEM 417
 FT TRANSMEM 418
 FT TRANSMEM 419
 FT TRANSMEM 420
 FT TRANSMEM 421
 FT TRANSMEM 422
 FT TRANSMEM 423
 FT TRANSMEM 424
 FT TRANSMEM 425
 FT TRANSMEM 426
 FT TRANSMEM 427
 FT TRANSMEM 428
 FT TRANSMEM 429
 FT TRANSMEM 430
 FT TRANSMEM 431
 FT TRANSMEM 432
 FT TRANSMEM 433
 FT TRANSMEM 434
 FT TRANSMEM 435
 FT TRANSMEM 436
 FT TRANSMEM 437
 FT TRANSMEM 438
 FT TRANSMEM 439
 FT TRANSMEM 440
 FT TRANSMEM 441
 FT TRANSMEM 442
 FT TRANSMEM 443
 FT TRANSMEM 444
 FT TRANSMEM 445
 FT TRANSMEM 446
 FT TRANSMEM 447
 FT TRANSMEM 448
 FT TRANSMEM 449
 FT TRANSMEM 450
 FT TRANSMEM 451
 FT TRANSMEM 452
 FT TRANSMEM 453
 FT TRANSMEM 454
 FT TRANSMEM 455
 FT TRANSMEM 456
 FT TRANSMEM 457
 FT TRANSMEM 458
 FT TRANSMEM 459
 FT TRANSMEM 460
 FT TRANSMEM 461
 FT TRANSMEM 462
 FT TRANSMEM 463
 FT TRANSMEM 464
 FT TRANSMEM 465
 FT TRANSMEM 466
 FT TRANSMEM 467
 FT TRANSMEM 468
 FT TRANSMEM 469
 FT TRANSMEM 470
 FT TRANSMEM 471
 FT TRANSMEM 472
 FT TRANSMEM 473
 FT TRANSMEM 474
 FT TRANSMEM 475
 FT TRANSMEM 476
 FT TRANSMEM 477
 FT TRANSMEM 478
 FT TRANSMEM 479
 FT TRANSMEM 480
 FT TRANSMEM 481
 FT TRANSMEM 482
 FT TRANSMEM 483
 FT TRANSMEM 484
 FT TRANSMEM 485
 FT TRANSMEM 486
 FT TRANSMEM 487
 FT TRANSMEM 488
 FT TRANSMEM 489
 FT TRANSMEM 490
 FT TRANSMEM 491
 FT TRANSMEM 492
 FT TRANSMEM 493
 FT TRANSMEM 494
 FT TRANSMEM 495
 FT TRANSMEM 496
 FT TRANSMEM 497
 FT TRANSMEM 498
 FT TRANSMEM 499
 FT TRANSMEM 500
 FT TRANSMEM 501
 FT TRANSMEM 502
 FT TRANSMEM 503
 FT TRANSMEM 504
 FT TRANSMEM 505
 FT TRANSMEM 506
 FT TRANSMEM 507
 FT TRANSMEM 508
 FT TRANSMEM 509
 FT TRANSMEM 510
 FT TRANSMEM 511
 FT TRANSMEM 512
 FT TRANSMEM 513
 FT TRANSMEM 514
 FT TRANSMEM 515
 FT TRANSMEM 516
 FT TRANSMEM 517
 FT TRANSMEM 518
 FT TRANSMEM 519
 FT TRANSMEM 520
 FT TRANSMEM 521
 FT TRANSMEM 522
 FT TRANSMEM 523
 FT TRANSMEM 524
 FT TRANSMEM 525
 FT TRANSMEM 526
 FT TRANSMEM 527
 FT TRANSMEM 528
 FT TRANSMEM 529
 FT TRANSMEM 530
 FT TRANSMEM 531
 FT TRANSMEM 532
 FT TRANSMEM 533
 FT TRANSMEM 534
 FT TRANSMEM 535
 FT TRANSMEM 536
 FT TRANSMEM 537
 FT TRANSMEM 538
 FT TRANSMEM 539
 FT TRANSMEM 540
 FT TRANSMEM 541
 FT TRANSMEM 542
 FT TRANSMEM 543
 FT TRANSMEM 544
 FT TRANSMEM 545
 FT TRANSMEM 546
 FT TRANSMEM 547
 FT TRANSMEM 548
 FT TRANSMEM 549
 FT TRANSMEM 550
 FT TRANSMEM 551
 FT TRANSMEM 552
 FT TRANSMEM 553
 FT TRANSMEM 554
 FT TRANSMEM 555
 FT TRANSMEM 556
 FT TRANSMEM 557
 FT TRANSMEM 558
 FT TRANSMEM 559
 FT TRANSMEM 560
 FT TRANSMEM 561
 FT TRANSMEM 562
 FT TRANSMEM 563
 FT TRANSMEM 564
 FT TRANSMEM 565
 FT TRANSMEM 566
 FT TRANSMEM 567
 FT TRANSMEM 568
 FT TRANSMEM 569
 FT TRANSMEM 570
 FT TRANSMEM 571
 FT TRANSMEM 572
 FT TRANSMEM 573
 FT TRANSMEM 574
 FT TRANSMEM 575
 FT TRANSMEM 576
 FT TRANSMEM 577
 FT TRANSMEM 578
 FT TRANSMEM 579
 FT TRANSMEM 580
 FT TRANSMEM 581
 FT TRANSMEM 582
 FT TRANSMEM 583
 FT TRANSMEM 584
 FT TRANSMEM 585
 FT TRANSMEM 586
 FT TRANSMEM 587
 FT TRANSMEM 588
 FT TRANSMEM 589
 FT TRANSMEM 590
 FT TRANSMEM 591
 FT TRANSMEM 592
 FT TRANSMEM 593
 FT TRANSMEM 594
 FT TRANSMEM 595
 FT TRANSMEM 596
 FT TRANSMEM 597
 FT TRANSMEM 598
 FT TRANSMEM 599
 FT TRANSMEM 600
 FT TRANSMEM 601
 FT TRANSMEM 602
 FT TRANSMEM 603
 FT TRANSMEM 604
 FT TRANSMEM 605
 FT TRANSMEM 606
 FT TRANSMEM 607
 FT TRANSMEM 608
 FT TRANSMEM 609
 FT TRANSMEM 610
 FT TRANSMEM 611
 FT TRANSMEM 612
 FT TRANSMEM 613
 FT TRANSMEM 614
 FT TRANSMEM 615
 FT TRANSMEM 616
 FT TRANSMEM 617
 FT TRANSMEM 618
 FT TRANSMEM 619
 FT TRANSMEM 620
 FT TRANSMEM 621
 FT TRANSMEM 622
 FT TRANSMEM 623
 FT TRANSMEM 624
 FT TRANSMEM 625
 FT TRANSMEM 626
 FT TRANSMEM 627
 FT TRANSMEM 628
 FT TRANSMEM 629
 FT TRANSMEM 630
 FT TRANSMEM 631
 FT TRANSMEM 632
 FT TRANSMEM 633
 FT TRANSMEM 634
 FT TRANSMEM 635
 FT TRANSMEM 636
 FT TRANSMEM 637
 FT TRANSMEM 638
 FT TRANSMEM 639
 FT TRANSMEM 640
 FT TRANSMEM 641
 FT TRANSMEM 642
 FT TRANSMEM 643
 FT TRANSMEM 644
 FT TRANSMEM 645
 FT TRANSMEM 646
 FT TRANSMEM 647
 FT TRANSMEM 648
 FT TRANSMEM 649
 FT TRANSMEM 650
 FT TRANSMEM 651
 FT TRANSMEM 652
 FT TRANSMEM 653
 FT TRANSMEM 654
 FT TRANSMEM 655
 FT TRANSMEM 656
 FT TRANSMEM 657
 FT TRANSMEM 658
 FT TRANSMEM 659
 FT TRANSMEM 660
 FT TRANSMEM 661
 FT TRANSMEM 662
 FT TRANSMEM 663
 FT TRANSMEM 664
 FT TRANSMEM 665
 FT TRANSMEM 666
 FT TRANSMEM 667
 FT TRANSMEM 668
 FT TRANSMEM 669
 FT TRANSMEM 670
 FT TRANSMEM 671
 FT TRANSMEM 672
 FT TRANSMEM 673
 FT TRANSMEM 674
 FT TRANSMEM 675
 FT TRANSMEM 676
 FT TRANSMEM 677
 FT TRANSMEM 678
 FT TRANSMEM 679
 FT TRANSMEM 680
 FT TRANSMEM 681
 FT TRANSMEM 682
 FT TRANSMEM 683
 FT TRANSMEM 684
 FT TRANSMEM 685
 FT TRANSMEM 686
 FT TRANSMEM 687
 FT TRANSMEM 688
 FT TRANSMEM 689
 FT TRANSMEM 690
 FT TRANSMEM 691
 FT TRANSMEM 692
 FT TRANSMEM 693
 FT TRANSMEM 694
 FT TRANSMEM 695
 FT TRANSMEM 696
 FT TRANSMEM 697
 FT TRANSMEM 698
 FT TRANSMEM 699
 FT TRANSMEM 700
 FT TRANSMEM 701
 FT TRANSMEM 702
 FT TRANSMEM 703
 FT TRANSMEM 704
 FT TRANSMEM 705
 FT TRANSMEM 706
 FT TRANSMEM 707
 FT TRANSMEM 708
 FT TRANSMEM 709
 FT TRANSMEM 710
 FT TRANSMEM 711
 FT TRANSMEM 712
 FT TRANSMEM 713
 FT TRANSMEM 714
 FT TRANSMEM 715
 FT TRANSMEM 716
 FT TRANSMEM 717
 FT TRANSMEM 718
 FT TRANSMEM 719
 FT TRANSMEM 720
 FT TRANSMEM 721
 FT TRANSMEM 722
 FT TRANSMEM 723
 FT TRANSMEM 724
 FT TRANSMEM 725
 FT TRANSMEM 726
 FT TRANSMEM 727
 FT TRANSMEM 728
 FT TRANSMEM 729
 FT TRANSMEM 730
 FT TRANSMEM 731
 FT TRANSMEM 732
 FT TRANSMEM 733
 FT TRANSMEM 734
 FT TRANSMEM 735
 FT TRANSMEM 736
 FT TRANSMEM 737
 FT TRANSMEM 738
 FT TRANSMEM 739
 FT TRANSMEM 740
 FT TRANSMEM 741
 FT TRANSMEM 742
 FT TRANSMEM 743
 FT TRANSMEM 744
 FT TRANSMEM 745
 FT TRANSMEM 746
 FT TRANSMEM 747
 FT TRANSMEM 748
 FT TRANSMEM 749
 FT TRANSMEM 750
 FT TRANSMEM 751
 FT TRANSMEM 752
 FT TRANSMEM 753
 FT TRANSMEM 754
 FT TRANSMEM 755
 FT TRANSMEM 756
 FT TRANSMEM 757
 FT TRANSMEM 758
 FT TRANSMEM 759
 FT TRANSMEM 760
 FT TRANSMEM 761
 FT TRANSMEM 762
 FT TRANSMEM 763
 FT TRANSMEM 764
 FT TRANSMEM 765
 FT TRANSMEM 766
 FT TRANSMEM 767
 FT TRANSMEM 768
 FT TRANSMEM 769
 FT TRANSMEM 770
 FT TRANSMEM 771
 FT TRANSMEM 772
 FT TRANSMEM 773
 FT TRANSMEM 774
 FT TRANSMEM 775
 FT TRANSMEM 776
 FT TRANSMEM 777
 FT TRANSMEM 778
 FT TRANSMEM 779
 FT TRANSMEM 780
 FT TRANSMEM 781
 FT TRANSMEM 782
 FT TRANSMEM 783
 FT TRANSMEM 784
 FT TRANSMEM 785
 FT TRANSMEM 786
 FT TRANSMEM 787
 FT TRANSMEM 788
 FT TRANSMEM 789
 FT TRANSMEM 790
 FT TRANSMEM 791
 FT TRANSMEM 792
 FT TRANSMEM 793
 FT TRANSMEM 794
 FT TRANSMEM 795
 FT TRANSMEM 796
 FT TRANSMEM 797
 FT TRANSMEM 798
 FT TRANSMEM 799
 FT TRANSMEM 800
 FT TRANSMEM 801
 FT TRANSMEM 802
 FT TRANSMEM 803
 FT TRANSMEM 804
 FT TRANSMEM 805
 FT TRANSMEM 806
 FT TRANSMEM 807
 FT TRANSMEM 808
 FT TRANSMEM 809
 FT TRANSMEM 810
 FT TRANSMEM 811
 FT TRANSMEM 812
 FT TRANSMEM 813
 FT TRANSMEM 814
 FT TRANSMEM 815
 FT TRANSMEM 816
 FT TRANSMEM 817
 FT TRANSMEM 818
 FT TRANSMEM 819
 FT TRANSMEM 820
 FT TRANSMEM 821
 FT TRANSMEM 822
 FT TRANSMEM 823
 FT TRANSMEM 824
 FT TRANSMEM 825
 FT TRANSMEM 826
 FT TRANSMEM 827
 FT TRANSMEM 828
 FT TRANSMEM 829
 FT TRANSMEM 830
 FT TRANSMEM 831
 FT TRANSMEM 832
 FT TRANSMEM 833
 FT TRANSMEM 834
 FT TRANSMEM 835
 FT TRANSMEM 836
 FT TRANSMEM 837
 FT TRANSMEM 838
 FT TRANSMEM 839
 FT TRANSMEM 840
 FT TRANSMEM 841
 FT TRANSMEM 842
 FT TRANSMEM 843
 FT TRANSMEM 844
 FT TRANSMEM 845
 FT TRANSMEM 846
 FT TRANSMEM 847
 FT TRANSMEM 848
 FT TRANSMEM 849
 FT TRANSMEM 850
 FT TRANSMEM 851
 FT TRANSMEM 852
 FT TRANSMEM 853
 FT TRANSMEM 854
 FT TRANSMEM 855
 FT TRANSMEM 856
 FT TRANSMEM 857
 FT TRANSMEM 858
 FT TRANSMEM 859
 FT TRANSMEM 860
 FT TRANSMEM 861
 FT TRANSMEM 862
 FT TRANSMEM 863
 FT TRANSMEM 864
 FT TRANSMEM 865
 FT TRANSMEM 866
 FT TRANSMEM 867
 FT TRANSMEM 868
 FT TRANSMEM 869
 FT TRANSMEM 870
 FT TRANSMEM 871
 FT TRANSMEM 872
 FT TRANSMEM 873
 FT TRANSMEM 874
 FT TRANSMEM 875
 FT TRANSMEM 876
 FT TRANSMEM 877
 FT TRANSMEM 878
 FT TRANSMEM 879
 FT TRANSMEM 880
 FT TRANSMEM 881
 FT TRANSMEM 882
 FT TRANSMEM 883
 FT TRANSMEM 884
 FT TRANSMEM 885
 FT TRANSMEM 886
 FT TRANSMEM 887
 FT TRANSMEM 888
 FT TRANSMEM 889
 FT TRANSMEM 890
 FT TRANSMEM 891
 FT TRANSMEM 892
 FT TRANSMEM 893
 FT TRANSMEM 894
 FT TRANSMEM 895
 FT TRANSMEM 896
 FT TRANSMEM 897
 FT TRANSMEM 898
 FT TRANSMEM 899
 FT TRANSMEM 900
 FT TRANSMEM 901
 FT TRANSMEM 902
 FT TRANSMEM 903
 FT TRANSMEM 904
 FT TRANSMEM 905
 FT TRANSMEM 906
 FT TRANSMEM 907
 FT TRANSMEM 908
 FT TRANSMEM 909
 FT TRANSMEM 910
 FT TRANSMEM 911
 FT TRANSMEM 912
 FT TRANSMEM 913
 FT TRANSMEM 914
 FT TRANSMEM 915
 FT TRANSMEM 916
 FT TRANSMEM 917
 FT TRANSMEM 918
 FT TRANSMEM 919
 FT TRANSMEM 920
 FT TRANSMEM 921
 FT TRANSMEM 922
 FT TRANSMEM 923
 FT TRANSMEM 924
 FT TRANSMEM 925
 FT TRANSMEM 926
 FT TRANSMEM 927
 FT TRANSMEM 928
 FT TRANSMEM 929
 FT TRANSMEM 930
 FT TRANSMEM 931
 FT TRANSMEM 932
 FT TRANSMEM 933
 FT TRANSMEM 934
 FT TRANSMEM 935
 FT TRANSMEM 936
 FT TRANSMEM 937
 FT TRANSMEM 938
 FT TRANSMEM 939
 FT TRANSMEM 940
 FT TRANSMEM 941
 FT TRANSMEM 942
 FT TRANSMEM 943
 FT TRANSMEM 944
 FT TRANSMEM 945
 FT TRANSMEM 946
 FT TRANSMEM 947
 FT TRANSMEM 948
 FT TRANSMEM 949
 FT TRANSMEM 950
 FT TRANSMEM 951
 FT TRANSMEM 952
 FT TRANSMEM 953
 FT TRANSMEM 954
 FT TRANSMEM 955
 FT TRANSMEM 956
 FT TRANSMEM 957
 FT TRANSMEM 958
 FT TRANSMEM 959
 FT TRANSMEM 960
 FT TRANSMEM 961
 FT TRANSMEM 962
 FT TRANSMEM 963
 FT TRANSMEM 964
 FT TRANSMEM 965
 FT TRANSMEM 966
 FT TRANSMEM 967
 FT TRANSMEM 968
 FT TRANSMEM 969
 FT TRANSMEM 970
 FT TRANSMEM 971
 FT TRANSMEM 972
 FT TRANSMEM 973
 FT TRANSMEM 974
 FT TRANSMEM 975
 FT TRANSMEM 976
 FT TRANSMEM 977
 FT TRANSMEM 978
 FT TRANSMEM 979
 FT TRANSMEM 980
 FT TRANSMEM 981
 FT TRANSMEM 982
 FT TRANSMEM 983
 FT TRANSMEM 984
 FT TRANSMEM 985
 FT TRANSMEM 986
 FT TRANSMEM 987
 FT TRANSMEM 988
 FT TRANSMEM 989
 FT TRANSMEM 990
 FT TRANSMEM 991
 FT TRANSMEM 992
 FT TRANSMEM 993
 FT TRANSMEM 994
 FT TRANSMEM 995
 FT TRANSMEM 996
 FT TRANSMEM 997
 FT TRANSMEM 998
 FT TRANSMEM 999
 FT TRANSMEM 1000

FT TRANSMEM 118 139 3 (POTENTIAL).
 FT TRANSMEM 140 160 4 (POTENTIAL).
 FT TRANSMEM 161 180 4 (POTENTIAL).
 FT TRANSMEM 181 205 5 (POTENTIAL).
 FT TRANSMEM 206 228 6 (POTENTIAL).
 FT TRANSMEM 229 248 7 (POTENTIAL).
 FT TRANSMEM 249 270 8 (POTENTIAL).
 FT TRANSMEM 271 291 9 (POTENTIAL).
 FT TRANSMEM 292 312 7 (POTENTIAL).
 FT TRANSMEM 313 333 7 (POTENTIAL).
 FT TRANSMEM 334 353 7 (POTENTIAL).
 FT TRANSMEM 354 373 7 (POTENTIAL).
 FT TRANSMEM 374 393 7 (POTENTIAL).
 FT TRANSMEM 394 413 7 (POTENTIAL).
 FT TRANSMEM 414 433 7 (POTENTIAL).
 FT TRANSMEM 434 453 7 (POTENTIAL).
 FT TRANSMEM 454 473 7 (POTENTIAL).
 FT TRANSMEM 474 493 7 (POTENTIAL).
 FT TRANSMEM 494 513 7 (POTENTIAL).
 FT TRANSMEM 514 533 7 (POTENTIAL).
 FT TRANSMEM 534 553 7 (POTENTIAL).
 FT TRANSMEM 554 573 7 (POTENTIAL).
 FT TRANSMEM 574 593 7 (POTENTIAL).
 FT TRANSMEM 594 613 7 (POTENTIAL).
 FT TRANSMEM 614 633 7 (POTENTIAL).
 FT TRANSMEM 634 653 7 (POTENTIAL).
 FT TRANSMEM 654 673 7 (POTENTIAL).
 FT TRANSMEM 674 693 7 (POTENTIAL).
 FT TRANSMEM 694 713 7 (POTENTIAL).
 FT TRANSMEM 714 733 7 (POTENTIAL).
 FT TRANSMEM 734 753 7 (POTENTIAL).
 FT TRANSMEM 754 773 7 (POTENTIAL).
 FT TRANSMEM 774 793 7 (POTENTIAL).
 FT TRANSMEM 794 813 7 (POTENTIAL).
 FT TRANSMEM 814 833 7 (POTENTIAL).
 FT TRANSMEM 834 853 7 (POTENTIAL).
 FT TRANSMEM 854 873 7 (POTENTIAL).
 FT TRANSMEM 874 893 7 (POTENTIAL).
 FT TRANSMEM 894 913 7 (POTENTIAL).
 FT TRANSMEM 914 933 7 (POTENTIAL).
 FT TRANSMEM 934 953 7 (POTENTIAL).
 FT TRANSMEM 954 973 7 (POTENTIAL).
 FT TRANSMEM 974 993 7 (POTENTIAL).
 FT TRANSMEM 994 1013 7 (POTENTIAL).
 FT TRANSMEM 1014 1033 7 (POTENTIAL).
 FT TRANSMEM 1034 1053 7 (POTENTIAL).
 FT TRANSMEM 1054 1073 7 (POTENTIAL).
 FT TRANSMEM 1074 1093 7 (POTENTIAL).
 FT TRANSMEM 1094 1113 7 (POTENTIAL).
 FT TRANSMEM 1114 1

QY 834 ATTGCT-----CTGTGACATGACTTAACTGTTAAAAA 796
 Db 165 ILeuysLeuSerIleTTPGlyLeuSerLeuLeuAlaLeuProValLeuLeuPheArg 184
 QY 795 AAACAGTGCCTTCAGATTAAGAAATTAGGGCTCTAATAGATGTTAATGATTAATGT 736
 Db 185 ArghValIlyrSerSerAsnValSerProAla-----CysTyGluAsp----- 199
 QY 735 CCTTTTGTGAGACAGAGTCTGTCTTAAAAACAACAAAGAAATGAGAGTTCTA 676
 Db 200 -----MetClysnAsnThrAlaAsnTPArgMetLeu 210
 QY 675 CTAACTGCTGCTGCTCCAGAGGGATTAATGAGAT----- 637
 Db 211 LeuArgIleLeuProGlnSerPheGlyPheIleValProLeuLeuIleMetLeuPheCys 230
 QY 636 -----CSTAAGAGAAACAAACACCTGAGGAGGAGCAGCATCTTGATTT 592
 Db 231 TyrGlyPheThrLeuArgThrLeuPheIlysalnIleMetGlyGlnIlyshIleArgAlaMet 250
 QY 591 CAGATTAAGTACCAAAAAATTTAGCTATCTTTCTGACAGACATGCTCTCTCTGTT 532
 Db 251 ArgValIlePheValIleValIlePheLeuLeuCysTTP-LeuProTyrAsnLeuVal 270
 QY 531 AACACTTCA-----GTCATTCAAAAGCCTATTGCTAGA 498
 Db 270 IleuLeuAlaAspThrLeuMetArgThrGlnValIleGlnIlyThrCysGluArg 288
 RESULT 7
 EX5C_BUCAI STANDARD: PRT; 1070 AA.
 AC P57528:
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Exoexyribonuclease V gamma chain (Ec 3.1.11.5).
 GN RECC OR BU453.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 symbiotic bacterium).
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxID=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tokyo 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT Buchnera sp. APS.";
 RL Nature 407:81-86(2000).
 CC -1- FUNCTION: EXHIBITS A WIDE VARIETY OF CATALYTIC ACTIVITIES
 INCLUDING ATP-DEPENDENT EXONUCLEASE, ATP-STIMULATED ENDONUCLEASE,
 CC ATP-DEPENDENT HELICASE AND DNA-DEPENDENT ATPASE ACTIVITIES
 CC (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
 CC ATP) in either 5'- to 3'- or 3'- to 5'-direction to yield 5'-
 CC phosphooligonucleotides.
 CC -1- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD
 CC (BY SIMILARITY).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AP001119; BAB1351.1;
 KW Hydrolyase; Nuclease; Exonuclease; Endonuclease; Helicase; DNA repair;
 SO COMPLETE PROTEOME.
 SO SEQUENCE 1070 AA; 128578 MW; 8F644C8487981AF CRC64;
 Alignment Scores:

Pred. No.: 1.29 Length: 1070
 Score: 85.50 Matches: 78
 Percent Similarity: 38.52% Conservative: 63
 Best Local Similarity: 21.31% Mismatches: 102
 Query Match: 3.68% Indels: 123
 DB: 1 Gaps: 22
 US-09-830-244b-2 (1-1331) x EX5C_BUCAI (1-1070)
 QY 261 TTCCGATGATGCTTTGAGAGTACTTCTTTGACAGAAAT---TTCATTCGCTTGC 317
 Db 105 PheGluAsnGlySerIlysalnLysAspMetIleLysIlyshPheLysPheSerPheMet 124
 QY 318 ATTGCTAATTCCTCC-----CTTTAAGAGCATGATTTCTTCTTTGTCG 368
 Db 125 AlaSerIlePheLysIlystyryIleLeuTyrArgProGlnIlyAsn-----GlnTTP 142
 QY 369 GAATGTCCTCCATTGACATTTTCAGATCTTTGATGTCACATAATGCCATTATTTGTAATG 428
 Db 143 GlnIleGluLys---AsnIleSerIle-PheAspLysAsn-----GlnGlnTTPGlnI 159
 QY 429 CCGTATGTCGATATACAGCATAGTTAATAACTGTTACAGTAATCTTACACTTGATTT 488
 Db 159 eLysLeuTTP---MetGlu-IleIleHisAsnThrLysIlyshIleAsnGlnSerAsp- 177
 QY 489 TGCCTGACCTTACCAATAGCTTTTGATGACTGAAAGTGTAAACAGAAAGAGCAT 548
 Db 178 ---HisPheAlaAsnLeuPheTyrAsnIleGlnIlyshIleLysGluLysLys--- 194
 QY 549 GTCTGACAGAAAGCATAGCTAATTAATTTTGTGACTTATCTGAATCCAGATGCTGCT 608
 Db 195 ---IleLysLysLysTyr-----L 200
 QY 609 TCCCTGCGAGCTGTTCTTCTTACGATCTCATTTGAATCC-----CTCT 656
 Db 200 euproLysArgPhePheIleIleSerPheSerMetAsnProSerTyrIleLysIleP 220
 QY 657 GGGAGCACAGACAGATTAGTAGAAGCTCCCATTC---TTTGTGTTTCTTTTAAAGACAG 713
 Db 220 heGlnAsnIleSerIleTyrThrAspIleTyrIlePheLeuTyrIleThrProPheLysTyr 240
 QY 714 AGACTCTGTCTCAAAAAAAGACATTATTCATTATTAACATCTTATAGACCCCTAAT 773
 Db 240 snIlePhe-----Asp 244
 QY 774 TCTTATCTGAAGCAGCTGTTTCTTTTAAACAGTTAAGTACGATGTCACAGACAAA 833
 Db 244 heIleGlnAspAsnLysIlePhe-----ThrAsp----- 253
 QY 834 TATTTCGATCAGATTAAGTCTCCCTGTCACAGTAGCAATGTTGTTCAATAGGGAGAG 893
 Db 254 -----IleLysIleGlu 258
 QY 894 AAAACAGATTTTAAAGTACTTTTGGGAGACTGATTTAGTAATATAAACTGCT 953
 Db 258 yLysAsnIleLeuAsn-----AspSerLeuIleThrLeuTTPG 271
 QY 954 CTCCTTAAGAAAAAACCCTTCCACCTTTACTGTCGTCATTTATATCCCTTATGTC 1013
 Db 271 LysGlnTyr-GluLeu-----IleTyrThrPheTyrIleLeuLysSer 284
 QY 1014 AAAGTTAATTTATCTTATTTCTGATATGCTTT-----TATACCAAGAGCGCTTAT 1064
 Db 285 LysLysAsnLysValIleAsn-----CysPheLysLysAsnLysAsn-LysSerLeu 302
 QY 1065 CAGCCAGTTCGAGAACACCATATACGACATACCATACCATACCATACCATACCATAC 1110
 Db 302 uSerGlnIleLysAsnAsnPheAsnAspSerGlnPheThrArgLysLysArgPheLe 322
 QY 1111 -----AATACCAATATATGTTTATATTCATCATCATCATCATCATCATCATCAT 1163
 Db 322 uGlnIleSerAspHisSerIleSerIleAsnIleCysPheAsnLysLysAsn----- 339

OY 1164 GAGGCTTGATGGCAATACGACTTCTACATCATATTCATCTTTCATACCAATAC 1223
 Db 340 -----GluilegluValLeuTyr-----G1 346
 OY 1224 ACTACTACACTTTTGTGATGATCATCTAAGACGACGATGAAACCCATTAATT 1283
 Db 346 ulyleuLeuLeuLeuLeu-----AsnGluAsnSerSerIleLysProGlyAspI1 363
 OY 1284 ACTGATACCT 1293
 Db 363 evalvalThr 366

RESULT 8

STAT_MACFA STANDARD: PRT: 61 AA.
 AC P02809; 21-JUL-1986 (Rel. 01, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE 15-DEC-1998 (Rel. 37, Last annotation update)
 GN Statherin precursor.

OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 NCBI_TaxID=9541;
 RN [1]
 RP PARTIAL SEQUENCE OF 1-19.
 RX MEDLINE=87309161; PubMed=3476566;
 RA Oppenheim F.G., Hay D.I., Smith D.J., Offner G.D., Troxler R.F.;
 RT "Molecular basis of salivary proline-rich protein and peptide
 RT synthesis: cell-free translations and processing of human and macaque
 RT statherin mRNAs and partial amino acid sequence of their signal
 RT peptides.";
 RL J. Dent. Res. 66:462-466(1987).
 RN [2]
 RP SEQUENCE OF 20-61.
 RX MEDLINE=82265555; PubMed=7107568;
 RA Oppenheim F.G., Offner G.D., Troxler R.F.;
 RT "Phosphoproteins in the parotid saliva from the subhuman primate
 RT Macaca fascicularis. Isolation and characterization of a proline-rich
 RT phosphoglycoprotein and the complete covalent structure of a proline-
 RT rich phosphopeptide.";
 RL J. Biol. Chem. 257:9271-9282(1982).
 RN [3]
 RP FUNCTION: SALIVARY PROTEIN THAT STABILIZES SALIVA SUPERSATURATED
 CC WITH CALCIUM SALTS BY INHIBITING THE PRECIPITATION OF CALCIUM
 CC PHOSPHATE SALTS. IT ALSO MODULATES HYDROXYAPATITE CRYSTAL
 CC FORMATION ON THE TOOTH SURFACE.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: SECRETED BY PAROTID AND SUBMANDIBULAR
 CC GLANDS.
 DR PIR: A03289; SBMOP1.
 KM Saliva: Phosphorylation; Enamel pellicle; Signal.

FT SIGNAL 1 19
 FT CHAIN 20 61
 FT MOD_RES 21 21
 FT MOD_RES 22 22
 FT DOMAIN 20 25
 FT DOMAIN 37 61
 FT SEQUENCE 61 AA; 7452 MW; 64241AAB5641ASB CRC64;
 SQ

Alignment Scores:

Pred. No.: 1.03 Length: 61
 Score: 85.00 Matches: 15
 Percent Similarity: 68.18% Conservative: 0
 Best Local Similarity: 68.18% Mismatches: 7
 Query Match: 3.66% Indels: 0
 DB: 1 Gaps: 0

US-09-830-244B-2 (1-1331) x STAT_MACFA (1-61)

OY 1059 CCTATACGCGAGTTCAGAAACCACTATACGCAACCACTACCAACCACTACCA 1118
 Db 40 ProtyrGlnProPheAlaProGlnProLeuTyrProGlnProTyrGlnPro 59
 OY 1119 CAATAT 1124
 Db 60 GlnTyr 61

RESULT 9

VO21_FOWPV STANDARD: PRT: 320 AA.
 ID VO21_FOWPV
 AC 093510;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 GN G-protein coupled receptor homology FPV021.
 OS FPV021.
 OS Fowlpox virus (FPV).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Avipoxvirus.
 NCBI_TaxID=10261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20193820; PubMed=10729156;
 RA Alonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
 RT "The genome of fowlpox virus.";
 RL J. Virol. 74:3815-3831(2000).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (Probable).
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: AF198100; AAF4365.1;
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR Pfam: PF00001; 7tm1.1.
 DR PRINTS: PR00237; GPCR_RHODOPSN.
 DR PROSITE: PS00337; G-PROTEIN_REC_P1_1; FALSE_NEG.
 DR PROSITE: PS00262; G-PROTEIN_REC_P1_2; 1.
 KM G-protein coupled receptor; transmembrane; Glycoprotein.
 FT DOMAIN 1 18
 FT TRANSMEM 19 39
 FT DOMAIN 40 52
 FT TRANSMEM 53 73
 FT DOMAIN 74 91
 FT TRANSMEM 92 112
 FT DOMAIN 113 133
 FT TRANSMEM 134 154
 FT DOMAIN 155 188
 FT TRANSMEM 189 209
 FT DOMAIN 210 222
 FT TRANSMEM 223 243
 FT DOMAIN 244 260
 FT TRANSMEM 261 281
 FT DOMAIN 282 320
 FT DISULFID 89 167
 FT CARBOHYD 175
 FT SEQUENCE 320 AA; 37808 MW; 5F082B9A1C6161958 CRC64;
 SQ

Alignment Scores:

Pred. No.: 1.26 Length: 320
 Score: 85.00 Matches: 56
 Percent Similarity: 39.51% Conservative: 40
 Best Local Similarity: 23.05% Mismatches: 59
 Query Match: 3.66% Indels: 88
 DB: 1 Gaps: 14

US-09-830-244B-2 (1-1331) x VO21_FOWPV (1-320)


```

RL NucleiAcids Symp. Ser. 31:233-234 (1994).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commerc
CC entities requires a license agreement (See http://www.isb-sib.ch/announc
CC or send an email to license@isb-sib.ch).
DR EMBL: D31785; BAA06573.1.
DR InterPro: IPR001750; Oxidored_g1.
DR Pfam: PF00361; Oxidored_g1.
DR KX Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
S0 SEQUENCE 567 AA; 66958 MW; 68C0F09BBF8D12BE CMC64;

Alignment Scores:
Score: 1.35 Length: 567
Percent Similarity: 85.00 Matches: 77
Best Local Similarity: 41.33% Conservative: 66
Query Match: 22.25% Mismatches: 128
DB: 3.66% Indels: 75
Gaps: 15

US-09-830-244B-2 (1-131) x NU2M_HANNMT (1-567)

QY 387 TTTTCAGATCTTTTGATGTCACATAATCCATTATGTGTAATGCCGATTATG---GTCAAT 44
DB ||| :|||: |||:|: :|||:|:|:| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 87 PheasnmetyrmetrPheLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 10
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 444 ACAGATGTTTAATTAACCTGTTTACCTAAATCTGATTTGCTGCACCTCTACC 50
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 504 AATACCTTTTGAATGACTGAAGTGTAAACAGAGAAGAGCCGTCTCCAGAAAGCA 56
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 117 -----AsnAsnAsnAsnAsnAsnMetSerMetGlyLys 12
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 564 TAGCTAATATTT---TTGCTACTTATCTGAATTCAGATGCTGCTCTCCCTGCACGT 62
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 127 TASPAsnTyrMetGlyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 14
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 621 TGTTCCTCTTACAGTCCTCATTTAAATCCCTCTGGAGACAGACGACTTACGAA 68
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 147 YLeuMetLeuLeuLeu---ThrSerAsnAsnLeuMet-----SerMetPheLeuSerMetG 16
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 681 CTCTCCATTTCTTTGTTTGTTTTAAACAGACAGACTCTG-----TCTCAAAAAAAG 73
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 165 LuLeuGlnSerTyrSerLeuTyrMetLeuThrGlyLeuMetProLysSerGlnLysSerG 18
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 735 GACATTTATATATATAACATCTTATT-----AGAGCCCTTAATTTCTTATCTGAAG 78
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 185 LysAsnSerLeuPheTyrTyrLeuMetGlyLysMetGlySerMetMetLeuTyrG 205
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 786 GCACTGTTTTTTTTTAAACAGTTAAGACGATGATGATCAACAGACCAATATTCTGTCA 84
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 205 LysMetSerLeuLeuTyrTyrMetThrSerAsnMetPheLeuAsnAsnLeuLeuLeu 224
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 846 GATAGTCCCTGTCAACAGTACGAATGTGGTTTCATTAAGTGGAGAAAGAACGACTTT 90
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 225 -----TyrSerLeuAsnMetTyrAsnLeuAsnAsnAsnMetLeu 238
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 906 TAAAGTAACTTTTGGAGAGACGATTATAGTAATATAATAAATC-----949
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 238 etGlyTyrPhePhe---IleIleMetGlyLeuLeuIleLysMetGlyAlaAlaProMet 257
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 950 -----TGTCTCCCTTAAAGAAAAAACCTTCACCTTACGAGTGCTATATATGCC 100
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 257 TyrAsnTyrPheMetLeuLeuTyrSerAsnSerPheAsnAsnMetMetMetMetMetMet 257
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


QY 1005 CTTAGTCCAAAGTATATCTTCTGATATATGCTTTATACCAAGACCTTAT 1064
 Db 277 -LeuMetProlysmetSerMetLeu-----SerTyrIleLeuIleMetLe 292
 QY 1065 CAGCCAGTCCGAGAACACACCTATACGACACACATACCAACCAATACCAATAT 1124
 Db 292 uasnleuTyraasnleuasnleuasnleuasnleuasnleuasnleuasnle 312
 QY 1125 ACCTTTT-----AATATCATCATGATACGACGACATGATTTAGACCTT----- 1170
 Db 332 uMetTyrMetLeuSerIleIleIle-----IleLeuSerIleuMetMetG 327
 QY 1171 -----GATTGGCAATACGACTTCTATCATCATCTTCT----- 1203
 Db 327 ySerMetGlyGlyLeuThrGlnMetLeuIleLysasnMetLeuLeuTyrSerGlyLeu 347
 QY 1204 -CATCTTCATACCATATACACATCTACTGACCTTTTGTNAGATCATTAAGACCAATGC 1262
 Db 347 uasnMetGlyTyrPheLeu-----LeuMetMetLeuSerIleuIleasnleu 363
 QY 1263 GAATGTAAACCCCTATATTTATGATGATCTTGTGTCAGATGATCTTGTCTTCCAA 1322
 Db 363 nasnleu-AsnSerMetLeuLeuLeuTyrMetTyrMetThrGlyTyrCysPheasnHis 383
 QY 1323 TGTCACTT 1330
 Db 383 etSerMet 385
 RESULT 11
 EDG3_HUMAN
 ID EDG3_HUMAN STANDARD: PRT: 378 AA.
 AC 099500:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lysosphingolipid receptor (EDG-3).
 GN EDG3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Placenta:
 RX MEDLINE=97032811; PubMed=8878560;
 RA Yamaguchi F., Tokuda M., Hatake O., Brenner S.;
 RT "Molecular cloning of the novel human G protein-coupled receptor
 (GPCR) gene mapped on chromosome 9.";
 RT Biochem. Biophys. Res. Commun. 227:608-614 (1996).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain:
 RX MEDLINE=98072391; PubMed=9409733;
 RA An S., Bleu T., Huang W., Hallmark O.G., Coughlin S.R., Goetzl E.J.;
 RT "Identification of cDNAs encoding two G protein-coupled receptors for
 lysosphingolipids.";
 RT FEBS Lett. 417:279-282 (1997).
 RL FEBS Lett. 417:279-282 (1997).
 CC -1- FUNCTION: ORPHAN RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES, BUT MOST ABUNDANTLY
 CC IN HEART, PLACENTA, KIDNEY, AND LIVER.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: X83864; CAA58744.1; -

DR EMBL: AF022139; AAC51906.1; -
 DR Genbank: HGNC:3167; EDG3.
 DR MIM: 601965; -
 DR InterPro: IPR004062; EDG3receptor.
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR InterPro: IPR004061; S1Preceptor.
 DR Pfam: PF00001; 7tm_1; 1.
 DR PRINTS: PR01524; EDG3RECEPTOR.
 DR PRINTS: PR00237; GPCR RHODOPSIN.
 DR PRINTS: PR01523; S1PRECEPTOR.
 DR PROSITE: PS00237; G-PROTEIN_RECEP_FL_1; 1.
 DR PROSITE: PS00237; G-PROTEIN_RECEP_FL_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.
 FT DOMAIN 1
 FT TRANSMEM 41 65
 FT DOMAIN 66 72
 FT TRANSMEM 73 101
 FT DOMAIN 102 115
 FT TRANSMEM 116 134
 FT DOMAIN 135 153
 FT TRANSMEM 154 179
 FT DOMAIN 180 195
 FT TRANSMEM 196 216
 FT DOMAIN 217 243
 FT TRANSMEM 244 265
 FT DOMAIN 266 281
 FT TRANSMEM 282 302
 FT DOMAIN 303 318
 FT CARBOHYD 15
 SQ SEQUENCE 378 AA; 42294 MW; 79A00306203F439F CMC64;
 Alignment Scores:
 Pred. No.: 1.63 Length: 378
 Score: 84.00 Matches: 42
 Percent Similarity: 42.35% Conservative: 41
 Best Local Similarity: 20.92% Mismatches: 69
 Query Match: 3.62% Indels: 44
 DB: 1 Gaps: 9
 US-09-830-244B-2 (1-1331) x EDG3_HUMAN (1-378)
 QY 653 CTCTGGGAGCAGACAGACAGTATAGAACCTCATCTTCTTGT----- 697
 Db 64 IleTyrPysAsnAsnLysPheHisAsnArgMetLysPheHeIleGlyAsnLeu 83
 QY 698 -----TTGTTTAAAGACAGACCTCTCTCAAAAAGAGACAT 739
 Db 84 CysAspLeuLeuAlaGlyIleAlaTyrLysValAsnIleLeuMetSerGlyLysTyr 103
 QY 740 TTATCATTAATACATCTTATTAGAGCCCTAAT-----TTCTATCTGAAGCAGCTTT 793
 Db 104 PheSerLeu-----SerProThrValTyrPheLeuArgGlySerMet 118
 QY 794 TTTTATTAAGACAGTAACTACT-----GATGTCAAGACGAATAA 835
 Db 119 PheValAlaLeuGlyAlaSerThrCysSerLeuLeuAlaIleAlaIleGlyArgHisLeu 138
 QY 836 TTTCTGATCAGATAGCCCTGTACACAGTAA----- 871
 Db 139 ThrMetLeuMetArgProTyrAspAlaAsnLysArgHisArgValPheLeuLeu 158
 QY 872 -----TGTGGTTCATTAAGTGGGAAGAAACAGATTAAAGTAACCTTTTGGGAGA 925
 Db 159 GlyMetCysTyrPheLeuLeuAlaPheThrLeuGlyAlaLeu--ProIleLeuGlyTyrPAsn 178
 QY 926 CT-----GATTGAGTAATAATAACTGTGTCTCCCTTAAGAAAAACCCCTTC 979
 Db 178 yLeuHisAsnLeuProAspCysSerThrIleLeuProLeuTyrSerLys--TyrI 197
 QY 980 ACCTTACGTGTC--ATTATATCCCTTGTGTCGAAGTAATATTATTATTTCTGC 1036
 Db 197 LeuIlePheCysIleSerIlePheThrAlaIleLeuValThrIleValIleLeuTyrAlaA 217

SQ SEQUENCE 442 AA: 49704 MW: 628D04DB20170A4B CRC64;
 Alignment Scores:
 Pred. No.: 2.11 Length: 442
 Score: 83.00 Matches: 47
 Percent Similarity: 39.81% Conservative: 35
 Best Local Similarity: 22.82% Mismatches: 76
 Query Match: 3.57% Indels: 48
 DB: 1 Gaps: 8
 US-09-830-244B-2 (1-1331) x TBB_TRYBR (1-442)
 QY 386 TTTTCACATCTTTGATGTCACATTAATGTAATGCGTTATGTTAGGATAAC 445
 DB 159 TyrProaspArgIleMetLeuThrPheSerIleLeuSerProlyValSerAspThr 178
 QY 446 AGCATAGTAAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 481
 DB 179 ValValIleuProlyrAsnThrThrLeuSerValHisIleuValGluAsnSerAspGlu 198
 QY 482 -----TTGGATTTGTCGACCTCTACCATTAAGCTTTGTAATGTAATGTAATGTAAT 532
 DB 199 SerMetCysIleAspAsnGluAlaLeuIleuThrPheSerIleLeuSerProlyValSerAspThr 218
 QY 533 ACAGAGAAAGAGCATGTCGACAGAAAGAGCATGTAATATTTTGTGTAATGTAATGTAAT 592
 DB 219 ThrProThrPheGlyAspLeuAsnHisLeuValSerAlaVal-----ValSer 234
 QY 593 AATCCAGATGTCGTC-----TTCCCTGAGGTTTTCCTGTCGTAATGTC----- 640
 DB 235 GlyValThrCysCysLeuArgPheProGlyGlnLeuAsnSerAspLeuArgIleVal 254
 QY 641 ----TCAITGATCCCTCTGGAGACAGACAGATAGTAGAACCTCATTTCTTGT 697
 DB 255 ValAsnLeuValProPhe-----ProArgLeuHisPheMet 267
 QY 698 TTGTTTAAAG-----ACAGAGACTGCTCTCAAAAAAGACATTTATCATTAATA 751
 DB 268 MetGlyPheAlaProLeuThrSerArgIleGlnIleuValGlyLeuSerValPro 287
 QY 752 CATCTTATTT-----AGAGCC 766
 DB 288 GluLeuThrGlnGlnMetPheAspAlaLysAsnMetGlnAlaAlaAspProArgHis 307
 QY 767 CCTATTCTTATTCGAGGACCTCTTTTAAACAGTAAGTACTGATGTCAC 826
 DB 308 GlyArgTyrLeuThrAlaSerAlaLeuPheArgIleGlyArgMetSerThrLysGluValAsp 327
 QY 827 AGACAATATTTTCATGATCAGATAGTCCCTGCAACAGTACCAATGCTTATCATTAAG 886
 DB 328 GluGlnMetLeuAsnValGln-----AsnLysAsnSerSerTyrPheIleGlu 343
 QY 887 TGGAGAGAAACACACATT 904
 DB 344 TrpIleProAsnAsnIle 349
 RESULT 15
 ID N120_YEAST STANDARD: PRT: 1037 AA.
 AC P35729: P35730:
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nucleoporin NUP120 (Nuclear pore protein NUP120).
 GN NUP120 OR RAT2 OR YKL057C OR YKL314 OR YKL313.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes;
 CC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RX MEDLINE=94378723; PubMed=8091862;

RA Rasmussen S.W.;
 RT "Sequence of a 28.6 kb region of yeast chromosome XI includes the
 RT FPA1 and TOA2 genes, an open reading frame (ORF) similar to a
 RT translationally controlled tumour protein, one ORF containing motifs
 RT also found in plant storage proteins and 13 ORFs with weak or no
 RT homology to known proteins.";
 RL Yeast 10:563-568(1994).
 RN (2)
 RP REVISIONS.
 RA Rasmussen S.W.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP CHARACTERIZATION, AND SEQUENCE OF 189-206 AND 800-807.
 RX MEDLINE=96134019; PubMed=8557736;
 RA Altchison J.D., Blobel G., Rout M.P.;
 RT "Nup120p: a yeast nucleoporin required for NPC distribution and mRNA
 RT transport.";
 RL J. Cell Biol. 131:1659-1676(1995).
 RN (4)
 RP PARTIAL SEQUENCE OF 550-555; 799-803 AND 375-384.
 RX MEDLINE=96152656; PubMed=8565072;
 RA Siniosoglou S., Wimmer C., Rieger M., Doye V., Tekotte H., Weise C.,
 RA Emig S., Segref A., Hurt E.C.;
 RT "A novel complex of nucleoporins, which includes Sec13p and a Sec13p
 RT homolog, is essential for normal nuclear pores.";
 RL Cell 84:265-275(1996).
 CC -1- FUNCTION: REQUIRED FOR EFFICIENT MRNA EXPORT FROM THE NUCLEUS TO
 CC -1- SUBUNIT: INTERACTS WITH NUP84, NUP85, SEC13 AND SEH1.
 CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 DR EMBL: X75781; CA653415.1; ALT_SEQ.
 DR EMBL: X75781; CA653414.1; ALT_SEQ.
 DR EMBL: Z28057; CA681894.1;
 DR PIR: S37879; S37879.
 DR SGD: S0001540; NUP120.
 KW Nuclear protein; Transport; Coiled coil.
 FT DOMAIN 435 465
 FT SIMILAR 434 763 TO NUP133 (AA 98-420).
 FT CONFLICT 384 384 K -> R (IN REF.4).
 FT SEQUENCE 1037 AA; 120447 MW; D4655E6116C54503 CRC64;
 Alignment Scores:
 Pred. No.: 2.34 Length: 1037
 Score: 83.00 Matches: 78
 Percent Similarity: 39.48% Conservative: 29
 Best Local Similarity: 28.78% Mismatches: 81
 Query Match: 3.57% Indels: 83
 DB: 1 Gaps: 19
 US-09-830-244B-2 (1-1331) x N120_YEAST (1-1037)
 QY 463 TGTATC-----AGTAATCTACACTTGGATTTGTCGACCTTACCAATAGCCTT 513
 DB 69 CysTyrHisPheSerArgSerThrLeuLeuThrPheTyrProLeuSerAspAlaTyr 88
 QY 514 -----TGAATGACTGAAGTGTATACACAGAAAGAGCATG 549
 DB 89 HisGlyLysThrIleAsnIleHisLeuProAsnAlaSerMetAsnGlnArgTyrThrLeu 108
 QY 550 TCTGCAGAAAGAGTACTTAATATTTTGGTA-----CTTATCTGAATCCAGATGCTG 606
 DB 109 ThrIleGlnValGlnGlnGlnLeuValAsnValIleLeuLysAspGlySerPhe 128
 QY 607 CTTCCTCGAGGTG---TTTCCTTCTTACGATCTCATTAATCCCTC----- 655

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 24, 2003, 19:16:12 ; Search time 90.5 Seconds

(without alignments)
6060.745 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 2322
Sequence: 1 ctatgttttagatcaaga.....gcctttccatgtaactgt 1331

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delpop 6.0 , Delpext 7.0	

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL-frame+ n2p.model -DEV-xlp
-Q/cgn2.1/USPTO.spool/US09830244/runat.24062003.130001.10042/app.query.fasta.1.1479
-DB-SPREMBL.21 -OEFMT-fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX-BLOSUM62 -TRANS-human40.cdl
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE-LOCAL -OUTFMT=ptc -NOR-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-NO_MAP -LANGQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMESOUT=120 -WARN_TIMESOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

SPREMBL.21:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_phage:
10: sp_plant:
11: sp_proteint:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:
15: sp_virus:
16: sp_bacteriap:
17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100.5	4.3	512	8	Q9XMS7 tetrahymena

Result No.	Score	Query Match	Length	ID	Description
2	96	4.1	387	5	Q9BRC6
3	92.5	4.0	673	10	Q9XG31
4	91.5	4.0	949	8	Q9BRX5
5	91	3.9	387	5	Q9BRC5
6	91	3.9	2708	5	Q15791
7	91	3.9	2742	5	Q15801
8	91	3.9	2819	5	Q15792
9	90	3.9	1560	5	Q9GRX5
10	89.5	3.9	307	10	Q9AVZ0
11	89	3.8	135	15	Q90814
12	88.5	3.8	474	2	Q9ZHL1
13	88.5	3.8	589	2	Q93Q09
14	88	3.8	1005	10	Q98RM3
15	87.5	3.8	332	5	Q9XUN3
16	87	3.8	227	5	Q8T1W4
17	87	3.8	855	10	Q9ZS83
18	87	3.8	1144	5	Q9G019
19	86.5	3.7	876	15	Q79670
20	86	3.7	443	5	Q25316
21	86	3.7	2028	8	Q32064
22	85.5	3.7	627	12	Q9EMS5
23	85.5	3.7	3933	5	Q97239
24	85	3.7	310	11	Q8VGW4
25	85	3.7	573	5	Q77334
26	85	3.7	577	10	Q9SMW8
27	85	3.7	620	8	Q9Z2P7
28	85	3.7	672	5	Q963G8
29	85	3.7	2166	12	Q36635
30	85	3.7	2391	5	Q27732
31	85	3.7	3124	5	Q9GVH8
32	84.5	3.7	311	5	Q966K6
33	84.5	3.7	2313	8	Q8WHM9
34	84	3.6	340	17	Q97A98
35	84	3.6	429	12	Q9EN15
36	84	3.6	651	10	Q9AM16
37	84	3.6	696	4	Q9G053
38	84	3.6	748	3	Q9N2N1
39	84	3.6	748	3	Q12311
40	84	3.6	1152	5	Q8T216
41	83.5	3.6	283	13	Q8UVA6
42	83.5	3.6	314	11	Q8VGY1
43	83.5	3.6	331	5	Q9U2Y0
44	83.5	3.6	515	17	Q97WM1
45	83.5	3.6	1844	5	Q97287

ALIGNMENTS

RESULT 1
ID Q9XMS7 PRELIMINARY: PRT; 512 AA.

AC Q9XMS7; 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Haem lyase.
CN YEJR.
OS Tetrahymena pyriformis.
OG Mitochondrion.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymena; Tetrahymena.
OX NCBI_TaxID=5908;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20181865; PubMed=10715208;
RA Edgyst J., Burger G., Gray M.N.;
RT "Expression of mitochondrial protein-coding genes in Tetrahymena pyriformis."
RL J. Mol. Biol. 297:381-393(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20181865; PubMed=10715207;
RA Burger G., Zhu Y., Littlejohn T.G., Greenwood S.J., Schnare M.N.,

...GATATGCTTTATACCAAGAGCCTATCA 1066

Figure 1


```

Db      692 LeuValIleLeuLeuLysAsnPhelIleSerIleLeuPheLeuThr----- 706
QY      1085 TGGTGTGTCTGGAAGTGGCTGTAAGCGCTTTGGTATTAAGCAATATCCAGAAATAG 1026
Db      707 -----IleLysTyrAspLysIleLysIleLeuAsnSerAsnLys 719
QY      1025 ATAAATTAACCTTGG-----AACTAAGGCGATATTAATGACACAGTAAGAGGTGAGGCTT 970
Db      720 ---IleAsnPhetylLysAsnPhelIleLysValSerPheIleLeuLysAsnPh 728
QY      969 TTTTTCCTTAAGGAGACAGACAGT-----TTTATTATTACTCAATC 928
Db      739 PheGlnIleLeuAsnAspLeuSerIlePheLeuTyrIleLeuPheIleLeuAlaGlyIle 758
QY      927 AGTCATCCCAAAAGTAC---TTTAAATGCTGTTTCTTCC-----CACTTATG 880
Db      759 IleserLysSerPheLysIleLysLeuIlePheAsnSerIleIleLeuGlyPheLys 778
QY      879 AAACCAATTTGGTACTGTTGACAGGGGAC-----TATCTGATCAGAAATATTGT 829
Db      779 LysPheLysSerPheSerLeuAspSerLysLeuAlaGlyAlaPhePheLysAsnPhLeu 798
QY      828 CTGTTGACATCAGTACTTAACTGTTTAAAAAAAACAGTGCCTTCAGATAAGAAATTA 769
Db      799 ILeValSerSerIleIle-----LeuLysLysCysSerIleLeuSerIlePheLys--- 815
QY      768 GGGGCTCTAATAGATGTATTAATGATTAATGTCCTTTTTCGACAGACAGTCTGTC 709
Db      815 ----- 815
QY      708 TTAATAAAACAACAAAGAAATGAGAGTCTTACTACTGCTCTGCTCCAGAGGGA 649
Db      816 ---LysSerLeuPheLysAspPheLeuIleLeuLeuThrGlnIle----- 829
QY      648 TTCATGAGCATGCTGAAGAAGAAAC-----AACTGAGGAGGAGACGACT 601
Db      830 ILeIleLeuLysArgLysLysGlyLysPheTyrValLeuAsnLeuIleAsnLeuThrPhe 849
QY      600 CTGGATTCATGATTAAGTAAAGTACCAAAAATATTAGTACTTCTTTCGACAGATGCTCT 541
Db      850 LeuArgLeuAsnSerAsnGlnLysAsnIlePheLysLysPheLysLysLysIle 869
QY      540 TTCTGTTTAAACACTTTCAGTCTTCAAAAGGCTATGTTAGAGGTGACGAAATCCAG 481
Db      870 PheGlyIle-MetIle---IleIleHisArgHisIleValLysIleLeuLysAsnSerAs 888
QY      480 TGTAGATTTACTGTAACAGTTTATTAACTATGCTGATTCACCAATAAGGCATTACCA 421
Db      888 nLeuAspLeuLeuIleTyrLeuLysArgMetCysIleGlu----- 902
QY      420 ATTAATGGCATTTAGTGCACATCAAAAAGATCTGAAAATGCTAATGAGGACATTTCCCAAAA 361
Db      903 -----LeuTyrSerLeuGlyValAsnTyrLeuLysAsnHisTyrAsnGlnAs 918
QY      360 GGAAGAATCCCAATGGCTCTATAAGGAGCAATATAGC 321
Db      918 nAsnLysAsn-----LeuLysAlaIleTyrSer 927

RESULT 5
Q9BK5 PRELIMINARY: PRT: 387 AA.
AC 09BK5:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Beta-tubulin (Fragment).
OS Malawimonas jakobiformis.
OC Eukaryota; Malawimonadidae; Malawimonas.
OX NCBI_TaxID=136089;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC50310;
RX MEDLINE=21165323; PubMed=11264402;

```

```

RA Edgcomb V.P., Roger A.J., Simpson A.G.B., Kysela D.T., Sogin M.L.:
RT "Evolutionary relationships among 'jakobi' flagellates as indicated
RL by alpha- and beta-tubulin phylogenies."
RM Mol. Biol. Evol. 18:514-522(2001).
DR EMBL: AF267186; AKS7437.1; -.
DR InterPro: IPR000217; Tubulin.
DR InterPro: IPR003008; Tubulin_fcsz.
DR Pfam: PF00091; tubulin_1.
DR PRINTS: PR01161; TUBULIN.
DR PROSITE: PS00227; TUBULIN; UNKNOWN_1.
KW GTP-binding.
FT NON_TER 1
FT NON_TER 387
SQ SEQUENCE 387 AA; 42840 MW; 6E3A6AC159AF90 CRC64;

Alignment Scores:
Pred. No.: 0.612 Length: 387
Score: 91.00 Matches: 44
Percent Similarity: 41.26% Conservative: 41
Best Local Similarity: 21.36% Mismatches: 73
Query Match: 3.92% Indels: 48
DB: 5 Gaps: 8

US-09-830-244b-2 (1-1331) x Q9BK5 (1-387)
QY 386 TTTTCAGATCTTTGATGTCGACTAATGCCATTATGTAATGCCATTATGCGAATAC 445
Db 144 TyrProAspArgMetCysThrPheSerValValProSerProLysValSerAspThr 163
QY 446 AGCATAGTAAATTAACCTGTTACAGTAATCTACAC----- 481
Db 164 ValValGluProTyrAsnAlaThrLeuSerValHisGlnLeuValGluAsnAlaAspGlu 183
QY 482 -----TTGATTTGCTGCACCTCTACCAATAGCCTTAAATGACGGAAGCTTA 532
Db 184 ValPheCysValAspAsnGlnAlaLeuTyrAspIleCysPheArgThrLeuLysLeuThr 203
QY 533 ACAGAGAAGAAAGCAGATCTGCGAAGAAAGATGATAATTTTGGTACTTATCTGA 592
Db 204 ThrProThrTyrGlyAspLeuAsnHisLeuValSerAlaVal-----MetSer 219
QY 593 AATCCAGATGCTGCG-----TTCCTGAGAGGTTGTTTCTTTCATGATCC----- 640
Db 220 GlyValThrCysLeuAlaGlyPheProGlyGlnLeuAsnSerAspLeuArgLysLeuAla 239
QY 641 ---TCATGTAATCCCTCTGGGACAGACAGACAGTATGAACTCCATTCCTTCTT 697
Db 240 ValAsnLeuIleProPhe-----ProArgLeuHisPhePheMet 252
QY 698 TTTGTTTTTAAAG-----ACAGAGACTGCTGCAAAAAGACATTATCATTAATNA 751
Db 253 ValGlyPheLeuAlaProLeuThrSerArgGlySerGlnGlnTyrArgAlaLeuThrValPro 272
QY 752 CATCTTATT-----AGAGCC 766
Db 273 GluLeuValGlnGlnMetPheAspAlaLysAsnMetLeuAlaLysAspProArgHis 292
QY 767 CCTAATTTCTTATCTGAAGGACACTGTTTTTTTAAACAGTAAGTACGATGATCAAC 826
Db 293 GlyHisTyrLeuThrAlaAlaValAlaPheArgGlyArgMetSerThrLysGlnValAsp 312
QY 827 AGACAAATATTTCTGATGATAGTCCCTGTCACACAGTACAGTAAGTGGCTTCAATAG 886
Db 313 GluGlnMetLeuAsnValGln-----AsnLysAsnSerSerTyrPheValGln 328
QY 887 TGGGAAGAAACAGCATT 904
Db 329 TrpIleProAsnAsnVal 334

RESULT 6
ID 015791 PRELIMINARY: PRT: 2708 AA.
AC 015791;
RX MEDLINE=21165323; PubMed=11264402;

```



```

Db      687 sSerGIuTySerTyrLysThrGlyValIleasnIleHisAspLeuLeuAspTyrPh 107
QY      1032 TCTGATATGCTTTATATCAAAAGAGCCTTATCAGCCAG-----1071
Db      707 ePheAspLeuHisPheIleLysAspGluLeuLysCysGlnSprThrAsnGlyAsnIleLys 727
QY      1072 -----TTCAGAACCAACCACTATACGACACACCAATACCAACCAATACCAATAT 1124
Db      727 sPheAspGlnMetAsnAspHisValLysAsnAspHisThrAsnTyrTyrGlnAsnAsnAs 747
QY      1125 AGCTTTATATATCATCAGT 1143
Db      747 nIlePheAsnLeuLeuThr 753

RESULT 8
ID      015792      PRELIMINARY:      PRT: 2819 AA.
AC      015792:
DT      01-JAN-1998 (TREMBLrel. 05, Created)
DT      01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      CG2.
GN      CG2.
OS      Plasmodium falciparum.
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_TaxID=5833;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-HB3;
RX      MEDLINE=98054002; PubMed=9393853;
RA      Su X., Kirkman L.A., Fujioke H., Wellens T.E.;
RT      "Complex polymorphisms in an approximately kDa protein are linked to
RL      chloroquine-resistant P. falciparum in Southeast Asia and Africa.";
RN      Cell 91:593-603(1997).
[2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-HB3;
RX      MEDLINE=20446201; PubMed=10989140;
RA      Fiodock D.A., Nomura T., Cooper R.A., Su X.-z., Talley A.K.,
RT      Wellens T.E.;
RT      "Allelic modifications of the cq2 and cgl genes do not alter the
RL      chloroquine response of drug-resistant Plasmodium falciparum.";
RL      Mol. Biochem. Parasitol. 110:1-10(2000).
DR      EMBL: AF030693; AAC47854.1; -;
SO      SEQUENCE 2819 AA; 335122 MW; FBF3CCCCOAC78BD9 CRC64;

Alignment Scores:
Pred. No.:      0.669      length:      2819
Score:          91.00      Matches:      41
Percent Similarity: 43.85%      Conservative: 41
Best Local Similarity: 21.93%      Mismatches: 58
Query Match:      3.92%      Indels:      47
DB:              5      Gaps:      9

US-09-830-244b-2 (1-1331) x 015792 (1-2819)
QY      669 CACTTGTAGTACACTCTCCATCTTCTTTGTTTATTAAGACAGACTCTGTCAAA 728
Db      580 GlnValAlaPheMetLeuProPheSerTyrPheSerAsnLysLysLys-----Lys 596
QY      729 AAAAAGACATTTATCATTAATAACATCATTAATAGACCCCTAATTTCTTATGTAG--- 785
Db      597 LysLys-----AsnIleIlePheLysAsnIleLeuSerTyrTyrGluTyr 611
QY      786 GCACGTGTTTTTTTTTAAACAGTAACTA-----CTGATGTCAACAGCAATATATTTC 839
Db      612 ValIleCysPhePheMetLysLysLysIleLysAsnIleLeuSerTyrAsnProTyrTyr 631
QY      840 TGATTCAGATAGTCCCTGTCAACAGTACCAATGTGGTTTCATTAAGTGGGAAGAAACA 899
Db      632 -IleHisIle-----GluSerTyrTyrAsnLysIleGly-----642

```

```

QY      900 GCATTTAAAGTACTTTTGGGAGACTGATTG-----AGTAATAA 941
Db      643 ---TYrCysIleValAlaPhePheLeuPheSerIleLeuLysIlePheCysThrCysAspAs 661
QY      942 TAAACGTCTGGTCTCCCTTAAGAAAAAACCCTTCCACCT-----984
Db      661 PASPTHTTHTLeuLeuIleSerTyrGlyShiValTyrTyrLeuSerAspSerLeuGluLys 681
QY      985 -----TACTGTGTCATTTATATCCCTTAGTCCCAAGTTAATATCTTATT 1031
Db      681 sSerGIuTySerTyrLysThrGlyValIleasnIleHisAspLeuLeuAspTyrPh 701
QY      1032 TCTGATATGCTTTATATCAAAAGCCTTATCAGCCAG-----1071
Db      701 ePheAspLeuHisPheIleLysAspGluLeuLysCysGlnSprThrAsnGlyAsnIleLys 721
QY      1072 -----TTCAGAACCAACCACTATACGACACACCAATACCAACCAATACCAATAT 1124
Db      721 sPheAspGlnMetAsnAspHisValLysAsnAspHisThrAsnTyrTyrGlnAsnAsnAs 741
QY      1125 AGCTTTATATATCATCAGT 1143
Db      741 nIlePheAsnLeuLeuThr 747

RESULT 9
ID      09GRX5      PRELIMINARY:      PRT: 1560 AA.
AC      09GRX5:
DT      01-MAR-2001 (TREMBLrel. 16, Created)
DT      01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT      01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE      Hypothetical 175.1 kDa protein.
GN      GDT1.
OS      Dictyostelium discoideum (Slime mold).
OC      Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
OX      NCBI_TaxID=44689;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-AX2;
RA      Pimpke G., Salger K., Wetterauer B.W.;
RT      "Absence of complementation in nonallelic mutants of Dictyostelium
RL      discoideum with defects in the transition from growth to
RT      development.";
RL      Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-AX2;
RX      MEDLINE=20255223; PubMed=10793140;
RA      Zeng C., Anjard C., Riemann K., Konzok A., Nellen W.;
RT      "gdt1, a new signal transduction component for negative regulation of
RL      the growth-differentiation transition.";
RL      Mol. Biol. Cell 11:1631-1643(2000).
DR      EMBL: AJ279060; CAC093932.1; -;
DR      InterPro: IPR000719; Euk_pkinase.
DR      InterPro: IPR002290; Ser_thr_kinase.
DR      Pfam: PF00069; pkinase.1.
DR      ProDom: PD000001; Euk_pkinase.1.
DR      SMART: SM00220; S_TKc.1.
DR      SMART: SM00219; TyrKc.1.
DR      PROSITE: PSS0011; PROTEIN_KINASE_DOM.1.
DR      ATP-binding: Hypothetical protein; Transferase.
FW      CONFLICT 20      I -> V (IN REF. 2).
FW      CONFLICT 619      F -> S (IN REF. 2).
SO      SEQUENCE 1560 AA; 175089 MW; 883B5BF9AD9E81EA CRC64;

Alignment Scores:
Pred. No.:      0.838      length:      1560
Score:          90.00      Matches:      56
Percent Similarity: 39.27%      Conservative: 41
Best Local Similarity: 22.67%      Mismatches: 106
Query Match:      3.88%      Indels:      44
DB:              5      Gaps:      7

```

RT nucleomorph."; [Proc. Natl. Acad. Sci. U.S.A. 97:200-205\(2000\)](#).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21223349; PubMed=11323671;
RA Douglas S., Zauner S., Franholz M., Beaton P., Penny S., Deng L.,
RA Wu X., Reith M., Cavalier-Smith T., Møller U.;
RT "The highly reduced genome of an enslaved algal nucleus."; [Nature 410:1091-1096\(2001\)](#).
RL Nature 410:1091-1096(2001).
DR EMBL: A010592; CAC27081.1; -.
KW Hypothetical protein
SQ SEQUENCE 307 AA; 37661 MW; 37E20BF8A1E0151D CRC64;

Alignment Scores:	
Pred. NO.:	0.883
Score:	89.50
Percent Similarity:	39.648
Best Local Similarity:	23.088
Query Match:	3.87%
DB:	10
Length:	307
Matches:	39
Conservative:	28
Mismatches:	47
Indels:	55
Gaps:	8

927 AGCTCCCAAAAGTACTTAAAATGCTGTTTCTCCACTTATGAACACACATTTG 868
 60 -----GIYTYrIleasnAnIlePheasp ----- 67

807 TCGTTAAAAAACAACAGTCCCTCAGATAGAGAAATTAACGGCGCTCAATATAGATGTAT 748
 88 -----LysylsullepheueneuleuGluylsty 97
 747 AATGATAATATGCTCTTTTGTGAGACAGAGTCTCTTAAAAAC ----- 700

Db	98	-----TyrProGluPheAsnLysIleIleLeuLeuIleValAsnLeuPheValTyr	114
QY	639	-----AAACAAGCAAMTGGAGAGTTCTACTACGTGCTGCTCCACAGG	652
		:	
Db	115	AspAsnSerPheLysArgLysPheTyrLysPhePheIleSerLysPheCysTrpTyrArg	134
QY	651	GGATTCT-----AATGAGATCGCTAAGAGAAAC	622
Db	135	AsnIleIleIleThrPheLeuAspIleLysValSerLysIleGluLysLysValPheAsn	154

```

QY      621  AACCCGAG---GGGAGACACATCTTGGATTTCAGATAAGTACCAAAAATATTAGCT 565
      622  :::::  |||  ::::  |||:::|||||  ::|||
Db      155  SerIleuYsthrAnLysAsnTrYPheAsnPhelLulIeLysLeuArgLysIlePhemet 174
      156  ::::  |||  ::::  |||:::|||||  ::|||
QY      564  ATCTCTTTCGCG-----AGACAT 547
      565  |||  |||  |||  |||  |||
Db      175  IleuPheCysThrPhePheArgHis 183
      176  ::::  |||  ::::  |||:::|||||  ::|||

RESULT 11

```

ID	PRELIMINARY;	PRT;	135 AA.
090814			
090814			
01-NOV-1998 (TREMblrel_08, Created)			
01-NOV-1998 (TREMblrel_08, last sequence update)			
01-DEC-2001 (TREMblrel_19, last annotation update)			
Envelope protein (Fragment).			
ENV.			
Human immunodeficiency virus type 2.			
HIVuses; Retroid viruses; Retroviridae; Lentiviruses			

acc.-1001D-11/05;
[1]
RRN
SEQUENCE FROM N.A.
SERP

RC STRAIN-A:
 RA MEDLINE=97255645; PubMed=910092;
 RX Xiang Z., Ariyoshi K., Wilkins A., Dias F., Whittle H., Breuer J.,
 RT "HIV type 2 pathogenicity is not related to subtype in rural guinea
 bissau".
 RL AIDS Res. Hum. Retroviruses 13:501-505(1997).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-A:
 RX MEDLINE=98406190; PubMed=9733826;
 RA Grassly N., Xiang Z., Ariyoshi K., Aaby P., Jensen H., Dias F.,
 RT Van der Loeff, Whittle H., Breuer J.,
 RT "Mortality among human immunodeficiency virus type 2-positive
 villages in rural guinea-bissau is correlated with viral genotype".
 RL J. Virol. 72:7895-7899(1998).
 DR EMBL: AJ011256; CA009567.1;
 DR InterPro: IPR000777; GP120.
 DR Pfam: PF00516; GP120.1.
 KM AIDS: Coat protein; Glycoprotein.
 FT NON_TER
 FT NON_TER
 SQ SEQUENCE 135 AA; 15598 MW; 2A732E706BBE663B CRC64;

Alignment Scores:
 Pred. No.: 0.965 Length: 135
 Score: 89.00 Matches: 31
 Percent Similarity: 39.34% Conservative: 17
 Best Local Similarity: 25.41% Mismatches: 36
 Query Match: 3.83% Indels: 3
 DB: 15 Gaps: 5

US-09-830-244B-2 (1-1331) x 090814 (1-135)

QY 916 TTTTGGAGAGCTGATTGAGTAATATAAAGCTGCTGCCCTTAGAGAAAAAACCC 975
 DB : : : : :
 DB 15 TTTTTPHISSE-----LysAsnAsnLysThrIleIleSerLeuAsnLysTYrAsn 32
 QY 976 TTCACCTTACTGT-----GTCAATTATATCCCTTAGTTCGA 1014
 DB : : : : :
 DB 33 LeuThrLeuHisCysLysArgProGluAsnLysThrValAlaProIleThrLeuMetSer 52
 QY 1015 AACTTAATATCTATATTCGATATCTGATTATACCAAGACCTTATACACCGCTTC 1074
 DB : : : : :
 DB 53 GlyLeuValAlaPheHisSerGlnProIleAsnThrLysProArgGlnAlaGlnCysGlnPhe 72
 QY 1075 CAGAAC-----AACCATATAG 1092
 DB : : : : :
 DB 73 LysGlyLysGlnLysLysAlaMetLeuGluValLysGlnThrLeuValLysHisProLys 92
 QY 1093 CACACACATACCAACCAACCATATACCAATATACGTTTAT----- 1134
 DB : : : : :
 DB 93 TyrLysGlnThrAsn--AspThrAsnAsnIleHisPheThrAlaProGluLysGlySer 111
 QY 1135 -----ATCATCAGTAAGTCCAGAGCATGATTATGAGCTTGATTGG 1176
 DB : : : : :
 DB 112 AsnProLysValAlaTyrMetLeuThrAsnCysArgGlyLysPheLeuSerLeuGlnTyr 131
 QY 1177 CAATA 1182
 DB : : : : :
 DB 132 Aspleu 133

RESULT 12
 092H1
 ID 092H1 PRELIMINARY: PRT: 474 AA.
 AC 092H1:
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE LSPB (Fragment).
 GN LSPB.
 OS Haemophilus ducreyi.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.

OX NCBI_TaxID=730;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=35000.
 RX MEDLINE=99030326; PubMed=9811662;
 RA Ward C.K., Lumley S.R., Latimer J.L., Cope L.D., Hansen E.J.,
 RT "Haemophilus ducreyi secretes a filamentous hemagglutinin-like
 protein".
 RL J. Bacteriol. 180:6013-6022(1998).
 DR EMBL: AF057696; AAC79760.1;
 FT NON_TER
 FT NON_TER
 SQ SEQUENCE 474 AA; 53579 MW; 8796871FD446A213 CRC64;

Alignment Scores:
 Pred. No.: 1.16 Length: 474
 Score: 88.50 Matches: 42
 Percent Similarity: 38.46% Conservative: 33
 Best Local Similarity: 21.54% Mismatches: 71
 Query Match: 3.83% Indels: 49
 DB: 2 Gaps: 10

US-09-830-244B-2 (1-1331) x 092H1 (1-474)

QY 885 TTTTGAACACCATTTGCTACTGTCAGAGGAC---TATCTGATCAGAAATTTGT 829
 DB : : : : :
 DB 82 PheValLeuProAsn-----LeuSerGlyLysIlePheSerIleHisAspleuAsp 98
 QY 828 CTGTGACATCAGTACTTAACTGTTTAAAAAAAACA-----GTGCCCTTCAGAT 778
 DB : : : : :
 DB 99 GlnLeuValGlnValLeuAsnThrValAsnLysArgAlaGluIleLysValLeuAlaSer 118
 QY 777 AAGAAATTAGGGGCTCTATATAGATGTTATATGATAAATGCTTTTTCGACAGA 718
 DB : : : : :
 DB 119 LysAlaTyrGlyLys-----SerAsn 125
 QY 717 GTCCTGCTTAAAAAACAAGAAAGATGAGAGCTTCTACTAATGCTGCTGCTCC 658
 DB : : : : :
 DB 126 LeuAsnIleLeuThrGlnArgThrArgLysTrpProThr--ValThrLeuSerIleAsn 144
 QY 657 CAGAGGGATTCATAGAGATGCTAGAGAGAAACAACCTGCAGGGAGAGACGATCTT 598
 DB : : : : :
 DB 145 AsnSerGly-----LysLusAsnGlnGlnAsnGlyArgAsnGlnMet 158
 QY 597 GGATTTCGATTAAGTACCAAAAAATATAGCTATCTCTTTCGACACATGCCCTTTC 538
 DB : : : : :
 DB 159 ThrLeuAsnValSerTrpSerAspLeuGlnGlyThrAsn----- 171
 QY 537 TCTGTAACACTTCACTGATTCATAAAGGCTATGTAAGTGCAGCAATCCACTGT 478
 DB : : : : :
 DB 172 -----AspValIlePheSerPheLysThrGlyTyr-----ArgLeuTyrLysGluThrLys 187
 QY 477 AGATTTCGTGACAGTATTATTAACATAGCTGATTCACCAATTAAGCGATTAACCAATA 418
 DB : : : : :
 DB 188 LysAsnThrGlnGlnAsnTyrSerLeuSerTyrIleGlnProPheSerTyrTyrThrLeu 207
 QY 417 ATGCGATTAGGCCATCAACAAGATCTGAAGAATGCTAATGG----- 376
 DB : : : : :
 DB 208 -----GluIleLysAlaSerGlnSerAlaTyrAsnLysGlnLeuThrGlyPhe 223
 QY 375 ---ACATTCCCAACAAGGAAGAAATCAATGCGCTCATTA 334
 DB : : : : :
 DB 224 TyrThrTyrProSerSerGlyLysThrGlnThrAlaAsnIleLys 238

RESULT 13
 093009
 ID 093009 PRELIMINARY: PRT: 589 AA.
 AC 093009:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Large superantigen protein exporter.
 GN LSPB.
 OS Haemophilus ducreyi.


```
Db      333 snValAlatValLysArGllleYsValLys-----LysLysSerL 347
Oy      357 AAGAAATCCAAANGCCTCTATAAAGGAGAATATAGCAATGGACAGCATGAATAATT 298
        ||          ::::           :|||         ||||
Db      347 yLeuUellrlyLyslleSerLysglilleSerlIethrrPhelystyPheLysllel 367
Oy      297 CTGTCAAAGAAAGTACTTCTCAACAACATCAGGAAAACTGTTCCCAATTAGGCCTGG 238
        |||:::|||||         |||       ::|    |||
Db      367 lephelysLysstyPhe-----GlAsnLyslleLeuLeuSerleuPhelel 384
Oy      237 GGAGCCACATATAGC---CACATATCTCAGATACTGATATTGTTT-----AGCGT 180
        |||||         |||||:::   |||  |||||:::|    ||||
Db      384 leUemertleerLysHstLyrlle-AsPLys-lleValleUrlyLysLysasBerle 403
Oy      189 TAGGGGCCCTCAAT-----ATTCCCCTCAT-----CTGTAAACTTCACATT 148
        |||  |||         |||       |||     :|||
Db      403 uleuUillyrAsnllellelleAsnlnHscglyLaArgLyVallLeuasn--Asnle 422
Oy      147 GTTACTCACAAGTCTGATAAATTCAGGCACATGACTT-----AACATTCC 100
        |          |||:::         |||||:::         :|||
Db      422 uSerlysrProleuAspserLystyrATrGmsSerBsnlleTyrasPLysCyshistIleva 442
Oy      99 CTATGGCAGTGCAAAATTCAGACATTTGTAAACACTGTAAATTTGACGTTTGATACATG 40
        ::|  |||  |||:::  |||         |||:::  |||
Db      442 lpheserArGLystyryPheThrIlePheLys-----AsnTyryPhephe----- 457
Oy      39 AGACAGCTTACCg 27
        |||||
Db      458 -----SerleuPro 460
```

RESULT 15			
ID	Q9XUN3	PRELIMINARY:	PRT: 332 AA.
AC	Q9XUN3		
DT	01-NOV-1999	(TREMBLrel. 12, Created)	
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)	
DT	01-MAR-2002	(TREMBLrel. 20, Last annotation update)	
DE	T20B3.5 protein.		
GN	T20B3.5		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;		
OC	Rhabditidae; Pelodermineae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Mortimore B.J.;		
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.		
RP	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99069613; PubMed=9851916;		
RA	none;		
RT	"Genome sequence of the nematode C.elegans: A platform for		
RT	investigating biology."		
RL	Science 282:2012-2018(1998).		
RL	EMBL: Z81593; CAB04743.1; "		
DR	InterPro: IPR003003; 7TM_chemo2.		
DR	InterPro: IPR000168; 7TM_nematode.		
DR	Pfam: PF01604; 7tm_5; 1.		
SO	SEQUENCE 332 AA; 38162 MW; 50903258D5357D9F CRC64;		

Alignment Scores:	
Pred. No.:	1 47
Score:	87.50
Percent Similarity:	33.99%
Best Local Similarity:	20.92%
Query Match:	3.77%
DB:	5
Length:	332
Matches:	64
Conservative:	40
Mismatches:	87
Indels:	115
Gaps:	15

US-09-830-244B-2 (1-1331) x Q9XUN3 (1-332)

QY 529 GTTAAACAGAGAAAGAGCATGTCTGCAGAAAGAGATAGCTAATATTTTTTGG----- 579
||| ::::| |:::| | |
Db 89 ValLeuYsGluIleGlyAlaSerValGluAlaGlnVal---TyrPheValValLeuVal 1077

QY	580	-----GTACTTTAT	588
Db	108	GIYAlATrMetMetValAlAlIleSerIleValPheGluSnaArgLeuPheValLeuVal	127
QY	589	CTGAATCCAGAGATGCTG-----CTTCCCTCGCAGTGTGT	624
Db	128	MetThrAsnLysMetLeuIleLysPheAlaLeuProIleTyrIleIleHisTyrIlePhe	147
QY	625	-----TTCCCTTCTTAGCATCCTCATGTGAATCCCTCTGGAGACACAGCAGTACTAG	678
Db	148	ProThrIleValLeuProSerLeuValLysLeuPro-----	159
QY	679	AACCTCCATTTCTTGTGTTGTTTTTAAACAGAGACTCTGTCCAAAAAAGACAC	738
Db	160	-----AspGlnGlnThrGlyLysAlaAsn	167
QY	739	TTTATCATTTAATACATCTTATTAGACCCCTAATTTCTTATCGAAGGACATGTTTTTT	798
Db	168	PheLeuLeu-----	170
QY	799	TTTAAACAGTTAGTGTCTGTATGTCACACAGCAATATTCTGTATCGATAGTCCCTGT	858
Db	171	-----LysTyrGlyCys-----ValProProTyr	178
QY	859	CAACAGTAGCAAAATGTGCTTCATTAAGTGGAGAAACAGCATTTTAAAGTAAC	915
Db	179	ValAspLeuGluArgValPheTyrLeuIleIleThrLysArgTyrPheLeuIleThrCys	198
QY	916	-----TTTTCG-----GAGACTGATTTTG	933
Db	199	AlaValPheIleCysThrMetPheAlaGluValTTrPhePheAlaLeuValThrPaspArg	218
QY	934	AGTAAATAAATAAATCTGCTCCCTTAAGAAAAAATAACCCCTTCACCCCTTTC	987
Db	219	LeuLeuLysGlnMetThrLysThrMetSerGlnLysThrPheAspLeuHisLysLys	238
QY	988	-----TGTGTCATTTTATTCGCCCTTAGTTCACAAAGTAATTAATCTTATTTCTGCAAT	1038
Db	239	PheGlnArgAlaPheIleLeuGlnLeuLeuIlePro---PheIleIleValPheLeuPro	257
QY	1039	ATTGCTTTTAAACCAAGCCCTTATTCAGCCACTTCACAGAACACACATATGCGCACAC	1098
Db	258	IleSerTyrIleGlyValAlaThrCysIleSerGlnTyrHisSnaIlnPhePhe-----	274
QY	1099	CATACACACCAATACCAATATATAGTTTTAATATCATCAGTAAGTACGAGACATGA	1158
Db	275	-----AsnAsnLeuThrLeuIleIleIleSerSerHisGly-----	286
QY	1159	TTATTAGGCTTGATTTGGCAATATACAGATTTCTATCCATATTTTCATATCCAT	1218
Db	287	PhePheSerThrIleAlaMetIleAlaLeuHisAlaPArgTyrArgIlnPheThrGlnLys	306
QY	1219	ATCAGACTACTACCATTTTGTGTTAGA---TCATTAACAGCAATGCGAATGAACACCC	1275
Db	307	Ile-----PheProCysLeuLysArgPheSerSerSerAlaThrSerSerThrAlaPro	324
QY	1276	TATATTTTACTGATACT	1293
Db	325	GlnSerLeuMetIleThr	330

Search completed: June 24, 2003, 19:24:00
Job time : 112.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 03:14:56 ; Search time 3443 Seconds

(Without alignments)
11250.605 Million cell updates/sec

Title: US-09-830-244B-2

Sequence: 1331
1 ccatgttttagaatcaag.....gcctttccaatgtcactg 1331

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Geneml:*

1: gb_ba:*

2: gb_hlg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hlg_hum:*

31: em_hlg_inv:*

32: em_hlg_other:*

33: em_hlg_mus:*

34: em_hlg_pln:*

35: em_hlg_rtd:*

36: em_hlg_mam:*

37: em_hlg_vrt:*

38: em_sy:*

39: em_hlg_hum:*

40: em_hlg_mus:*

41: em_hlg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	704	52.9	133984	9	AL512306	AL512306 Human DNA
C 2	396	29.8	153023	2	AC021462	AC021462 Homo sapi
C 3	361.2	27.1	153023	2	AC021462	AC021462 Homo sapi
C 4	349.2	26.2	146805	2	AL450424	AL450424 Homo sapi
C 5	253	19.0	542	9	HUMSTRNA	M18371 Human stath
C 6	253	19.0	552	9	HUMSTRNA	M18078 Human stath
C 7	253	19.0	1584	9	AK092678	AK092678 Homo sapi
C 8	176.2	13.2	297	11	G05711	G05711 human STS
C 9	145.4	10.9	1498	11	G26643	G26643 human STS
C 10	145.4	10.9	4723	9	HUMSTATH2	M32639 Human saliv
C 11	145.4	10.9	141568	9	AC063956	AC063956 Homo sapi
C 12	145.4	10.9	161549	2	AC024676	AC024676 Homo sapi
C 13	86.2	6.5	524	9	HUMHIS2X	M26665 Human hist
C 14	86.2	6.5	558	9	BC009791	BC009791 Homo sapi
C 15	81	6.1	480	9	HUMHIS1X	M26664 Human hist
C 16	81	6.1	566	9	BC017835	BC017835 Homo sapi
C 17	78.8	5.9	491	9	HUMHHRPA	M18372 Human hist
C 18	77.2	5.8	209410	2	AL805933	AL805933 Mus muscu
C 19	74.6	5.6	9881	6	AX281311	AX281311 Sequence
C 20	72	5.4	193445	2	AL845502	AL845502 Mus muscu
C 21	64.8	4.9	81962	2	AC128396	AC128396 Rattus no
C 22	64.8	4.9	246173	2	AC094026	AC094026 Rattus no
C 23	62.2	4.7	185969	2	AC069037	AC069037 Homo sapi
C 24	59	4.4	1764	9	HUMHIS103	L05512 Human hist
C 25	59	4.4	185969	2	AC069037	AC069037 Homo sapi
C 26	55.8	4.2	9881	6	AX281312	AX281312 Sequence
C 27	52.4	3.9	164399	3	PFMAL3P6	Z99551 Plasmodium
C 28	52.2	3.9	128635	2	AC108074	AC108074 Homo sapi
C 29	52.2	3.9	154472	2	AC109457	AC109457 Homo sapi
C 30	52.2	3.9	207558	2	AC114778	AC114778 Homo sapi
C 31	51.6	3.9	170141	2	AC092022	AC092022 Homo sapi
C 32	51.6	3.9	171044	2	AC092787	AC092787 Homo sapi
C 33	51.6	3.9	194874	2	AC080090	AC080090 Homo sapi
C 34	51.6	3.9	274349	2	AC093623	AC093623 Homo sapi
C 35	51.2	3.8	12929	3	AE001373	AE001373 Plasmodium
C 36	51	3.8	182362	9	AL160033	AL160033 Human DNA
C 37	48.6	3.7	1141	6	AX083744	AX083744 Sequence
C 38	48.6	3.7	157051	9	AC011095	AC011095 Homo sapi
C 39	48.6	3.7	167830	2	AC027768	AC027768 Homo sapi
C 40	48.6	3.7	173765	9	AC090457	AC090457 Homo sapi
C 41	48.4	3.6	6056	6	AX345928	AX345928 Sequence
C 42	48.4	3.6	121709	9	AC093783	AC093783 Homo sapi
C 43	48.4	3.6	129989	9	AC005681	AC005681 Homo sapi
C 44	48.4	3.6	174277	9	AC093826	AC093826 Homo sapi
C 45	48	3.6	61215	2	AC090357	AC090357 Homo sapi

ALIGNMENTS

RESULT 1

AL512306

LOCUS Human DNA sequence from clone Rpl1-430C7 on chromosome 1, complete

DEFINITION

AL512306

ACCESSION

AL512306

VERSION

AL512306.16 GI:18491332

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 133984)

AUTHORS

Bray-Allen,S.

TITLE

Direct Submission

JOURNAL

JOURNAL	COMMENT
Submitted (01-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humbly@esanger.ac.uk Clone requests: clonerequests@sanger.ac.uk	On Feb 4, 2002 this sequence version replaced at:18477311

RPL1-430C7 is a 1000 bp cDNA fragment constructed by the group of Pletzer de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6
 IMPORTANT: This sequence is not the entire insert of clone RPL1-430C7. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
 The true right end of clone RPL1-430C7 is at 13384 in this sequence. The true left end of clone RPL1-739N20 is at 97715 in this sequence. The true right end of clone RPL1-7317 is at 2000 in this sequence.

FEATURES

FEATURES	
source	location/qualifiers
	1. .133984
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="1"
	/clone="RP11-430C7"
	/clone_1id="RPC1-11.2"
misc.feature	123904. .123933
	/note="Single clone region. Reads generated from a transposon library derived from a single pUC clone. Restriction digest data confirm the assembly."
BASE COUNT	37368 a 32097 c 30202 g 34317 t
ORIGIN	

Query Match	Best Local Similarity	52.98%	Score 704	DB 9	Length 133984
Matches 718	Conservative	0	Mismatches	1.6e-140	Indels 1; Gaps 1;
QY	1	CTATGTTTTAGATCAAAAGATGAAACCGGTAAAGCTGTCTCATGTACCAAACTGGAATTT	60		
Db	92638	CTATGTTTTAGATCAAAAGATGAAACCGGTAAAGCTGTCTCATGTACCAAACTGGAATTT	9257		
QY	61	ACAGTGTTTACAAATGCTGTGAAATTTTTCACCTGCATAGGAAATGTTAAAGTACTTGCGC	120		
Db	92578	ACAGTGTTTACAAATGCTGTGAAATTTTTCACCTGCATAGGAAATGTTAAAGTACTTGCGC	9251		
QY	121	TGAAATTTATCACACTGTGAGTAAACAATTGAAAGTTTAGCAGATGAGGGCAATATTTG	180		
Db	92518	TGAAATTTATCACACTGTGAGTAAACAATTGAAAGTTTAGCAGATGAGGGCAATATTTG	9245		
QY	181	AGGCCCCCTAAAGCTAAACAATAATACAGATCTGAGATAGTGGCTAAATGGCTCCCA	240		
Db	92458	AGGCCCCCTAAAGCTAAACAATAATACAGATCTGAGATAGTGGCTAAATGGCTCCCA	9239		
QY	241	GGGCTAATTTGGACAAGTTTTCCGATGTGCTTTGAGAATACCTTCTTTTGCAGAA	300		

Db	92398	GGCCTAATTGGGGAACGTTTTCCGATGTGCTTAGAAGTACTTTTGGACAGAA	92339
QY	301	TTTTCAATTCGCTTGGCATTGCTAATATTCCTTTATAGAGCCATTGGATTTCTTTCC	360
Db	92338	TTTTCAATTCGCTTGGCATTGCTAATATTCCTTTATAGAGCCATTGGATTTCTTTCC	92279
QY	361	TTTTGTGGGAATGTCCATTACCATTTTCAGATCTTTTGAAGTGGACATAAGCATAT	420
Db	92278	TTTTGTGGGAATGTCCATTACCATTTTCAGATCTTTTGAAGTGGACATAAGCATAT	92219
QY	421	TGCTAATGCCGTTATTGGTGAATACAGCATAGTTAAATAACTGTTATACGTAATCTACA	480
Db	92218	TGGTAATGCCGTTATTGGTGAATACAGCATAGTTAAATAACTGTTATACGTAATCTACA	92159
QY	481	CTTGATTTTCTGCTACCTCTACCAATAGACCTTTTGAATACAGAAAGTTTAAACAGAA	540
Db	92158	CTTGATTTTCTGCTACCTCTACCAATAGACCTTTTGAATACAGAAAGTTTAAACAGAA	92099
QY	541	AGAGGCATGTCTGCAGAAAGAGATAGCTAATATTTTGGTACTTATCTGAAATCCAAAG	600
Db	92098	AGAGGCATGTCTGCAGAAAGAGATAGCTAATATTTTGGTACTTATCTGAAATCCAAAG	92039
QY	601	ATCTCTCTTCCCTCGAGGTGTGTTCCCTTTAGAGCCCATATGATCCCTCTGGGA	660
Db	92038	ATCTCTCTTCCCTCGAGGTGTGTTCCCTTTAGAGCCCATATGATATCCCTCTGGGA	91979
QY	661	GCAAGGACAGTTAGTAGAAGCTCCCATTTCTTGTGTTGTTTTTAAGACAGACACTCT	720
Db	91978	GCAAGGACAGTTAGTAGAAGCTCCCATTTCTTGTGTTGTTTTTAAGACAGACACTCT	91920
QY	721	GTCCT 724	
Db	91919	CTCT 91916	

RESULT 2

AC021462/c	AC021462	153023 bp	DNA	linear	HTG 03-APR-2000
LOCUS	AC021462				
DEFINITION	Homo sapiens clone RP11-276C1, WORKING DRAFT SEQUENCE, 21 unordered pieces				
ACCESSION	AC021462				
VERSION	AC021462.3	GI:7387343			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	1 (bases 1 to 153023)				
JOURNAL	Birren,B., Linton,L., Nusbaum,C. and Lander,E.				
REFERENCE	Homo sapiens, clone RP11-276C1				
AUTHORS	Unpublished				
	2 (bases 1 to 153023)				
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F., Boguslavsky,L., Bonkshgalter,B., Brown,A., Burkett,G., Castle,A., Chappel,T., Colangelo,M., Collins,M., Collymore,A., Cooke,P., DeBellano,K., Dewar,K., Domini,M., Doyle,M., Fenesor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L., Howard,J.C., Johnson,R., Jones,C., Kann,L., Kartas,A., Klein,J., Lander,T., Lechoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGuck,A., McKernan,K., McPherson,R., Meldrum,T., O'Donnell,L., Morrow,J., Naylor,T., Norman,C.H., O'Connor,T., Womansell,P., Oliver,T.M., Peterson,K., Plierre,N., Pisanil,C., Pollard,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Rothman,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (JL-JUN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT	On Apr 3, 2000 this sequence version replaced at:7230200.				

*	NOTE: This is a 'working draft' sequence. It currently		
*	consists of 25 contigs. The true order of the pieces		
*	is not known and their order in this sequence record is		
*	arbitrary. Gaps between the contigs are represented as		
*	runs of N, but the exact sizes of the gaps are unknown.		
*	This record will be updated with the finished sequence		
*	as soon as it is available and the accession number will		
*	be preserved.		
*	1		
*	5105:	contig of 5105 bp	in length
*	5106	5205: gap of	100 bp
*	5206	10570: contig of 5365 bp	in length
*	10571	10670: gap of	100 bp
*	10671	17628: contig of 6956 bp	in length
*	17627	17726: gap of	100 bp
*	17727	20478: contig of 2752 bp	in length
*	20479	20578: gap of	100 bp
*	20579	22033: contig of 2457 bp	in length
*	23036	23135: gap of	100 bp
*	23136	25308: contig of 2173 bp	in length
*	25309	25408: gap of	100 bp
*	25409	30465: contig of 5061 bp	in length
*	30470	30569: gap of	100 bp
*	30570	32688: contig of 2119 bp	in length
*	32689	32788: gap of	100 bp
*	32789	36187: contig of 3399 bp	in length
*	36188	36287: gap of	100 bp
*	36288	48988: contig of 12701 bp	in length
*	48989	49088: gap of	100 bp
*	49089	67333: contig of 18245 bp	in length
*	67334	67433: gap of	100 bp
*	67434	82317: contig of 14684 bp	in length
*	82318	82417: gap of	100 bp
*	82418	84614: contig of 2197 bp	in length
*	84615	84714: gap of	100 bp
*	84715	90336: contig of 5612 bp	in length
*	90337	90426: gap of	100 bp
*	90427	92647: contig of 2221 bp	in length
*	92648	92747: gap of	100 bp
*	92748	97690: contig of 4943 bp	in length
*	97691	97790: gap of	100 bp
*	97791	100799: contig of 3009 bp	in length
*	100800	100899: gap of	100 bp
*	100900	104202: contig of 3303 bp	in length
*	104203	104302: gap of	100 bp
*	104303	107125: contig of 2855 bp	in length
*	107158	107257: gap of	100 bp

FEATURES
 source
 * 107258 110300: contig of 3043 bp in length
 * 110301 110400: gap of 100 bp
 * 110401 113955: contig of 3555 bp in length
 * 113956 114055: gap of 100 bp
 * 114056 118801: contig of 4746 bp in length
 * 118802 118901: gap of 100 bp
 * 118902 129931: contig of 10930 bp in length
 * 129932 137545: gap of 100 bp
 * 137546 137645: gap of 100 bp in length
 * 137646 146805: contig of 9160 bp in length.
 Location/Qualifiers
 1. 146805
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone_1b="RPCI-11.2"
 1. 5105
 /note="assembly-fragment:00368
 clone_end:T7
 vector_side:left"
 5206. 10570
 /note="assembly-fragment:00594
 fragment_chain:1"
 10671. 17626
 /note="assembly-fragment:00786
 fragment_chain:1"
 17727. 20478
 /note="assembly-fragment:00844
 fragment_chain:1"
 20579. 23035
 /note="assembly-fragment:00011
 fragment_chain:1"
 23136. 25308
 /note="assembly-fragment:01029
 fragment_chain:1"
 25409. 30469
 /note="assembly-fragment:00148
 fragment_chain:2"
 30570. 32688
 /note="assembly-fragment:00961
 fragment_chain:2"
 32789. 36187
 /note="assembly-fragment:00310
 fragment_chain:3"
 36288. 48988
 /note="assembly-fragment:01150
 fragment_chain:3"
 49089. 67333
 /note="assembly-fragment:00444
 fragment_chain:4"
 67434. 82317
 /note="assembly-fragment:00140
 fragment_chain:4"
 82418. 84614
 /note="assembly-fragment:00657
 fragment_chain:5"
 84715. 90326
 /note="assembly-fragment:00757
 fragment_chain:5"
 90427. 92647
 /note="assembly-fragment:00677
 fragment_chain:6"
 92748. 97690
 /note="assembly-fragment:00248
 fragment_chain:6"
 97781. 100799
 /note="assembly-fragment:00091"
 100900. 104202
 /note="assembly-fragment:00538"
 104303. 107157
 /note="assembly-fragment:00675"

misc_feature 107258..110300
 /note="assembly-fragment:00695"
 misc_feature 110401..113955
 /note="assembly-fragment:01316"
 misc_feature 114056..118801
 /note="assembly-fragment:01292
 fragment_chain:7"
 misc_feature 118902..129931
 /note="assembly-fragment:00551
 fragment_chain:7"
 misc_feature 129932..137545
 /note="assembly-fragment:01410
 fragment_chain:7"
 misc_feature 137646..146805
 /note="assembly-fragment:00763
 fragment_chain:7"
 clone_end:SP6
 vector_side:right"
 BASE COUNT 37508 a 33240 c 34076 g 39557 t 2424 others
 ORIGIN

Query Match 26.2% Score 349.2; DB 2: Length 146805;
 Best Local Similarity 96.3%; Pred. No. 7.9e-65;
 Matches 368; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 709 GACAGAGACTCTGTCTCAAAAAAAGACATTATCATTAATATTAAGAGCCCC 768
 DB 84740 GACAGAGACTCTGTCTCAAAAAAAGACATTATCATTAATATTAAGAGCCCC 84799
 QY 769 TAATTTCTATCTGAAGCAGCTGTTTTTTTAAACAGTAATCTGATGATCAACG 828
 DB 84800 TAATTTCTATCTGAAGCAGCTGTTTTTTTAAACAGTAATCTGATGATCAACG 84859
 QY 829 ACAATATTTCTGATCAGATAGTCCCTGTCAACAGTACAAATGCTTCATTAAGTG 888
 DB 84860 ACAATATTTCTGATCAGATAGTCCCTGTCAACAGTACAAATGCTTCATTAAGTG 84919
 QY 889 GGAAGAAAACAGCATTTTAAAGTAATCTTTGGGAGACTGATTGAGTAATATAAACT 948
 DB 84920 GGAAGAAAACAGCATTTTAAAGTAATCTTTGGGAGACTGATTGAGTAATATAAACT 84978
 QY 949 CTGGTCTCCTTAAGAAAAAACCCTTCCACCTTCTGTCATTTATATCCCTTA 1008
 DB 84979 CTGGTCTCCTTAAAGAAAAAACCCTTCCACCTTCTGTCATTTATATCCCTTA 85038
 QY 1009 GTTCCAAGTTAATATCTTATTTCTGATATTTGTTTATACCAAGAGCTTATCAGC 1068
 DB 85039 GTTCCAAGTTAATATCTTATTTCTGATATTTGTTTATACCAAGAGCTTATCAGC 85098
 QY 1069 CAGTTCAGAACACACTATA 1090
 DB 85099 CAGTTCAGAACACTATA 85120

RESULT 5
 HOMOSTRNA
 LOCUS 542 bp mRNA linear PRI 13-JAN-1995
 DEFINITION Human statherin mRNA, complete cds.
 ACCESSION M18371
 VERSION M18371.1 GI:338610
 KEYWORDS statherin.
 SOURCE Human female submandibular gland, cDNA to mRNA, clone pBRHSP9B8.2.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 542)
 AUTHORS Dickinson,D.P., Ridall,A.L. and Levine,M.J.
 TITLE Human submandibular gland statherin and basic histidine-rich
 peptide are encoded by highly abundant mRNA's derived from a common
 ancestral sequence
 JOURNAL Biochem. Biophys. Res. Commun. 149 (2), 784-790 (1987)
 MEDLINE 88106506
 PUBMED 3426601

Magatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and Isogai, T.
 NEDO human cDNA sequencing project
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 2 (bases 1 to 1584)
 Isogai, T. and Yamamoto, J.
 Direct Submission
 Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team): 2-6-7
 Kazusa-Kamatari, Katsuzo, Chiba 292-0812, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology (RAB): cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.): 5'-6' and one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
 Location/Qualifiers

COMMENT

FEATURES

Source
 1. 1584
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="SALGI100107"
 /tissue_type="salivary gland"
 /clone_lib="SALGI1"
 /note="cloning vector: pME18SFL3"
 89..997
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAC03943.1"
 /db_xref="GI:21751328"
 /translation="MGDSEMLKLPVQKCEHKLKMKARLSGYEALAKTIFKTKDKSP
 EMSEKFLIKKFWTDSNAVVOGLAEALVYVNAHVAGTKGCVSVSVSVKFNQPK
 AKKEIGIEICLMYIEIEKGEAVOEELKLDNKNPKTIIVATIEFLKRALSFSGKIL
 LKPLIKVLKPLFESEKAVREDAKILAIETIRWIDALRPLOININSVOLKEEEM
 VKPTAPRPTRPLRSQLEKLEKLEQDSAGSDAGGDDGDEVPQIDAYELLENVEI
 LSKLPTEMTKLRQKNGKREKRPWSL"

CDS

BASE COUNT 526 a 298 c 331 g 429 t
 ORIGIN

Query Match 19.0%; Score 253; DB 9; Length 1584;
 Best Local Similarity 88.4%; Pred. No. 6,6e-44;
 Matches 274; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

1022 TATATCTTATTCGATATTCCTTTATACCAAGAGCCTTTCAGCCAGTTCCAGACA 1081
 1174 TTTCGCTAGATTTGGAAGATTGGTTATGGTATGCGCTTATCAGCCAGTTCCAGACA 1233
 1082 ACCACTATTCGACCAACCAATACCAACCAATACCAATATACGTTTATATATCA 1141
 1234 ACCACTATACCAACCAATACCAACCAATACCAATATACGTTTATATATCA 1293
 1142 GTAACTGACGACATGATTTAGGCTTGATGGCAATACGACTTTACATCCATAT 1201
 1294 GAACTGACGACATGATTTAGGCTTGATGGCAATACGACTTTACATCCATAT 1353
 1202 CTCATCTTCATACATATCACACTACTACCACTTTGTNAGATCATTAAGACATG 1261
 1354 CTCATCTTCATACATATCACACTACTACCACTTTGTNAGATCATTAAGACATG 1413
 1262 GGAATGTAACCCATATATTTAGGATCTTTGGTTCCAGATCTGCTTTTCA 1321
 1414 CAATGTAACCCATATATTTAGGATCTTTGGTTCCAGATCTGCTTTTCA 1473
 1322 ATGTCACCTG 1331
 1474 TTGTCACCTG 1483
 Db

RESULT 8
 G05711
 LOCUS 297 bp DNA linear STS 19-OCT-1995

DEFINITION
 human STS WI-7844, sequence tagged site.
 G05711
 ACCESSION
 VERSION
 G05711.1 GI:858956
 STS: STS sequence; primer: sequence tagged site.
 KEYWORDS
 SOURCE
 Homo sapiens STS derived from sequences in dbEST and the UniGene collection.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Whitehead Institute/MIT Center for Genome Research; Physically Mapped ESTs
 Unpublished (1995)

Contact: Thomas Hudson
 Whitehead Institute/MIT Center for Genome Research
 Whitehead Institute for Biomedical Research
 9 Cambridge Center, Cambridge MA 02142 USA
 Tel: 617 252 1900
 Fax: 617 252 1902
 Email: thudson@genome.wi.mit.edu

Primer A: ACTGACGACATGATTTATGACG
 Primer B: AATAATGCTTGAATTTATGAC
 STS size: 286
 PCR Profile:

Presoak:
 Denaturation: 56 degrees C
 Annealing: 56 degrees C
 Polymerization: 35
 PCR Cycles: 35
 Thermal Cycler:
 Protocol:
 Template: 10 ng
 Primer: each 5 pm
 dNTPs: each 4 mM
 Taq Polymerase: 0.025 units/ul
 Total Vol: 20 ul

Buffer:
 MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3

Prepared with primer pairs derived from M18371 -- Unigene.

FEATURES
 source
 Location/Qualifiers
 1..297
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="748_F_3: 793_E_2: 794_A_9: 886_B_3: 960_D_2:
 972_E_4: 750_D_11: 855_E_12: 882_G_11: 931_G_10: 395.8 CR
 from top of Chr4 linkage group"

STS
 primer_bind
 12..297
 12..34
 primer_bind
 99 a 56 c 37 g 105 t
 BASE COUNT
 ORIGIN
 complement(274..297)

Query Match 13.2%; Score 176.2; DB 11; Length 297;
 Best Local Similarity 92.9%; Pred. No. 2.2e-27;
 Matches 184; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

1134 TATCATGTAAGTACGACGACATGATTTAGGCTTGATGGCAATACGATTTACA 1193
 1 TATCATGTAAGTACGACGACATGATTTAGGCTTGATGGCAATACGATTTACA 60
 1194 TCCATATTCATCTTTATACCAATACGACACTACTTCTTTGTGATGATCATCTAA 1253
 61 TCCATATTCATCTTTATACCAATACGACACTACTTCTTTGTGATGATCATCTAA 120
 1254 GAGCAATGGAATGTAACCCATATTTACTGATACCTTTGGTTCCAGATACCTGC 1313
 Db

```

Db      121 GAGCAATGCAATGAAAAACACTATATTTACTGTATACCTTTGTTTCAGGATACCTTGC 180
OY      1314 CTTTTCAGTGTCACTTG 1331
Db      181 CTTTTCAGTGTCACTTG 198

RESULT 9
LOCUS   G26643
DEFINITION human STS STS_M32639, sequence tagged site.
ACCESSION G26643
VERSION  G26643.1 GI:1348875
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE   Homo sapiens STS derived from sequences in dbEST and the UniGene
          collection.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Hudson,T.
TITLE    Whitehead Institute/MIT Center for Genome Research; Physically
          Mapped STSs
JOURNAL  Unpublished (1995)
COMMENT  Contact: Thomas Hudson
          Whitehead Institute/MIT Center for Genome Research
          Whitehead Institute for Biomedical Research
          9 Cambridge Center, Cambridge MA 02142 USA
          Tel: 617 252 1900
          Fax: 617 252 1902
          Email: thudson@genome.wi.mit.edu
          Primer A: GGACATGATATTGTAGGTAAGATGG
          Primer B: ACTGCCTTCAACCATACACAGG
          STS size: 210
          PCR Profile:
          Presoak:
          Denaturation:
          Annealing: 56 degrees C
          Polymerization:
          PCR Cycles: 35
          Thermal Cycler:
          Protocol:
          Template: 10 ng
          Primer: each 5 pm
          dNTPs: each 4 mM
          Tag Polymerase: 0.025 units/ul
          Total Vol: 20 ul
          Buffer:
          MgCl2: 1.5 mM
          KCl: 50 mM
          Tris-HCL: 10 mM
          pH: 9.3
FEATURES
source   Derived from dbEST (genbank accession M32639).
          1..1498
          Location/Qualifiers
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /map="473.8 cr from top of Chr4 linkage group"
          STS
          primer_bind
          21..230
          primer_bind
          complement(209..230)
          BASE COUNT 511 a 229 c 244 g 514 t
          ORIGIN
          Query Match 10.9%; Score 145.4; DB 11; Length 1498;
          Best Local Similarity 90.1%; Pred. NO. 6.1e-21;
          Matches 155; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

OY      1160 TATTGAGCTTGATTTGGCAATACGACTTCTACATTCATATTCATCTTTCATACCATTA 1219
          1..1498
          Db      1234 TGTGACAGGCTTGTGATGGCAATACGACTTCTACATTCATATTCATCTTTCATACCATTA 1293
          1..1498
          OY      1220 TCACACTACACACCTTTTGTNAGATCATCTAAGACCAATGCAATGTAAACCCATATA 1279
          1..1498
          Db      1294 TCACACTACACACCTTTTGTNAGATCATCTAAGACCAATGCAATGTAAACCCATATA 1353
          1..1498
          OY      1280 ATTACTGATACCTTTGCTGCTCCAGATACCTTCTTTCATATTCATCTTG 1331
          1..1498
          Db      1354 ATTACTGATACCTTTGCTGCTCCAGATACCTTCTTTCATATTCATCTTG 1405
          1..1498

RESULT 10
LOCUS   HUMSTAT2
DEFINITION Human salivary statherin gene, exons 2-6.
ACCESSION M32639
VERSION  M32639.1 GI:338504
KEYWORDS statherin.
SEGMENT 2 of 2
SOURCE   Human (individuals #563, #8136, and J.F.) fibroblast, cell line
          #563, DNA, clones 1-3.
          Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Sabatini,L.M., He,Y.Z. and Azen,E.A.
TITLE    Structure and sequence determination of the gene encoding human
          salivary statherin
JOURNAL  Gene 89 (2), 245-251 (1990)
MEDLINE  90323623
PUBMED  2373369
COMMENT  Draft entry and computer-readable sequence for [1] kindly submitted
          by L.M.Sabatini, 03-JAN-1990.
FEATURES
source   1..4723
          Location/Qualifiers
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /map="4q11-q13"
          join(M31077.1:1601..1931,1..352)
          /gene="STATH"
          /note="STATH mRNA and introns"
          <1..286
          /gene="STATH"
          /note="STATH intron A"
          join(302..352,1532..1552,1645..1674,2728..2814)
          /note="statherin precursor"
          /codon_start=1
          /protein_id="AAA60593.1"
          /db_xref="GI:338506"
          /translation="MKELVFAFILALWVMIGADSSSEKFLRIGRGYGPYPVP
          EQPLYPQYPOPOOYTF"
          join(302..352,1532..1537)
          /note="statherin signal peptide"
          join(1538..1552,1645..1674,2728..2811)
          /product="statherin"
          <302..352
          /gene="STATH"
          /note="statherin precursor, (first translated exon);
          G00-120-391"
          /number=2
          353..1531
          /note="STATH intron B"
          1532..1552
          /number=3
          1553..1644
          /note="STATH intron C"
          1645..1674
          /number=4
          1675..2727
          /note="STATH intron D"

```

```

exon          2728..>2814
              /note="statherin precursor"
              /number=5
              2848..4051
              /note="STATH Intron E"
              4294..4299
BASE COUNT    1570 a 797 c 722 g 1634 t
ORIGIN        About 1.8 kb after segment 1.

Query Match   10.9% Score 145.4; DB 9; Length 4723;
Best Local Similarity 90.1%; Pred. NO. 4.8e-21;
Matches 155; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1160 TATTGAGCTTGGTGAATAGCAATGATTCATATTCATCTTCATACACATA 1219
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4045 TGTGACAGCTTGATGGCAATAGCAATGATTCATATTCATCTTCATACACATA 4104

QY 1220 TCACACTACTACACTTTTGTGATGATCATCTAGACCAATGCGATGTAACCCCTATA 1279
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4105 TCACACTACTACACTTTTGTGATGATCATCTAGACCAATGCGATGTAACCCCTATA 4164

QY 1280 ATTACTAGATGATCTTGTGATGATCATCTAGACCAATGCGATGTAACCCCTATA 1331
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4165 ATTACTAGATGATCTTGTGATGATCATCTAGACCAATGCGATGTAACCCCTATA 4216

RESULT 11
AC063956      141568 bp      DNA      linear      PRI 25-AUG-2000
LOCUS         Homo sapiens 4 BAC RP11-529K3 (Roswell Park Cancer Institute Human
AC063956      AC063956      complete sequence.
VERSION       AC063956.7 GI:9910030
KEYWORDS      HTG.
SOURCE        Homo sapiens.
ORGANISM      Homo sapiens.
REFERENCE     Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 141568)
              Muzny,D.M., Adams,C., Adio-Oduola,B., All-ouman,F.R., Allen,C.,
              Albrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barberia,J.,
              Benton,J., Blinze,K., Blankenburg,K., Bonnin,D., Bouck,J.,
              Bowler,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
              Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
              Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
              Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
              Coyne,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
              Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
              Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
              Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
              Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
              Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
              Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunatirne,P., Hale,S.,
              Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
              He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
              Holloway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Hully,S.,
              Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
              Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U.,
              Kling,L., Korvah,J., Kovar,C., Kratovic,Z., Kureshi,A., Landry,N.,
              Ling,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lien,C.,
              Liu,J., Liu,W., Louisedge,H., Lozano,R.J., Lu,X., Lucier,A.,
              Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R.,
              Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P.,
              Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T.,
              Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M.,
              Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
              Nguyen,N., Nickerson,E., Nnokenko,S., Oguh,M., Okunolu,G.,
              Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
              Peters,L., Pickens,R., Plimus,E., Pu,L.L., Quiles,M., Ren,Y.,
              Rieves,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
              Scherer,S., Scott,G., Shen,H., Shoohtari,N., Slsson,I.,
              Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H.,
              Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,

```

```

TITLE         Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
REFERENCE     Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wall,R.,
AUTHORS        Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S.,
TITLE         Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
JOURNAL       Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,
              Naylor,S.L. and Gibbs,R.
              Direct Submission
              Unpublished
              2 (bases 1 to 141568)
              Morley,K.C.
              Direct Submission
              Submitted (22-APR-2000) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
              3 (bases 1 to 141568)
              Morley,K.C.
              Direct Submission
              Submitted (25-AUG-2000) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
              On Aug 25, 2000 this sequence version replaced g1:9795448.
              INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
              gc-help@bcm.tmc.edu

```

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT

```

----- Summary Statistics -----
Contig Length: 141568
Phrap values in estimate: 140895
Average error rate (BCM-Phrap estimate): 1.4137e-05
Fraction of Phrap values less than 40 : 0.00715426
Number of consensus changing edits: 10
Number of N's in consensus : 0

```

```

----- Consensus changing edits -----
Position Original+Context Edited+Context
23165 cctgaattct(n)ttctgtgcaa cctgaattct(t)ttctgtgcaa

```


Db 126696 ATTACTGTATCTCTTTCTTTCAGAGTACTGCCCTTTTCATTTCTCACTTG 126747

|||||

RESULT 13
LOCUS HUMHIS2X 524 bp mRNA linear PRI 31-DEC-1994
DEFINITION Human histatin 2 (HIS2) mRNA, complete cds.
ACCESSION M26665
VERSION M26665.1 GI:292145
KEYWORDS histatin 2.
SOURCE Homo sapiens parotid gland cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 524)
AUTHORS Sabatini, L.M. and Acen, E.A.
TITLE Histatins, a family of salivary histidine-rich proteins, are encoded by at least two loci (HIS1 and HIS2)
JOURNAL Biochem. Res. Commun. 160 (2), 495-502 (1989)
MEDLINE 89246491
PUBMED 27196777

FEATURES
source location/Qualifiers
1..524 /organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="parotid gland"
1..524 /gene="HIS2"
72..227 /gene="HIS2"
/codon_start=1
/product="histatin 2"
/protein_id="AA58646.1"
/db_xref="GI:292146"
/translation="MKFFVALLALMLSMTGADSHAKRHGKRFHEKHSHRGYR
SNLYDN"
506..511 /gene="HIS2"
BASE COUNT 157 a 100 c 87 g 180 t
ORIGIN

Query Match 6.5%; Score 86.2; DB 9; Length 524;
Best Local Similarity 62.5%; Pred. No. 3.2e-08;
Matches 150; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

QY 1093 CACACCATACCAACCAATATACGTTTAATATCATCTAGTACGCG 1152
|||||
Db 187 CACATCGAGGCTATGATCAATATCTGTATGACATGATATCTTCAGTATCATGGG 246
1153 ACATGATTTATGAGGCTTGTGGAATATGACATTCATATTCATCTTCA 1212
|||||
Db 247 GCATGATTTATGAGGCTTGTGGAATATGACATTCATATTCATCTTCA 306
1213 TACCATATGACATCTACTACCTTTTGTGATCATCTTAAGCAATGC-GAATGTAA 1271
|||||
Db 307 TACCGATATGACATCTACTACCTTTTGTGATCATCTTAAGCAATGC-AATGAAG 366
1272 ACCCTAATATTTACTGATATCTTGTGTCAGATATCTGCTTTCATATGACCTTG 1331
|||||
Db 367 AATATACATGATTTAGTGAATCTGTGCTTTCAGATATCTGCTTTCATATGACCTTG 426

RESULT 14
LOCUS BC009791 558 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, histatin 3, clone MGC:13578 IMAGE:4293405, mRNA,
complete cds.
ACCESSION BC009791
VERSION BC009791.1 GI:14602560
KEYWORDS MGC.
SOURCE MGC. sapiens.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLOMTECH
CDNA Library Preparation: CLOMTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: amadanesystemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Keltman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 19 Row: 0 Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4557652.
Location/Qualifiers
1..558 /organism="Homo sapiens"
/db_xref="locusID:3347"
/db_xref="taxon:9606"
/clone="MGC:13578 IMAGE:4293405"
/tissue_type="Skeletal Muscle"
/clone_lib="NIH-MGC_81"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
76..231 /codon_start=1
/product="histatin 3"
/protein_id="AAH09791.1"
/db_xref="GI:14602561"
/translation="MKFFVALLALMLSMTGADSHAKRHGKRFHEKHSHRGYR
SNLYDN"
BASE COUNT 187 a 102 c 91 g 178 t
ORIGIN

Query Match 6.5%; Score 86.2; DB 9; Length 558;
Best Local Similarity 62.5%; Pred. No. 3.2e-08;
Matches 150; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

QY 1093 CACACCATACCAACCAATATACGTTTAATATCATCTAGTACGCG 1152
|||||
Db 191 CACATCGAGGCTATGATCAATATCTGTATGACATGATATCTTCAGTATCATGGG 250
1153 ACATGATTTATGAGGCTTGTGGAATATGACATTCATATTCATCTTCA 1212
|||||
Db 251 GCATGATTTATGAGGCTTGTGGAATATGACATTCATATTCATCTTCA 310
1213 TACCATATGACATCTACTACCTTTTGTGATCATCTTAAGCAATGC-GAATGTAA 1271
|||||
Db 311 TACCGATATGACATCTACTACCTTTTGTGATCATCTTAAGCAATGC-AATGAAG 370
1272 ACCCTAATATTTACTGATATCTTGTGTCAGATATCTGCTTTCATATGACCTTG 1331
|||||
Db 371 AATATACATGATTTAGTGAATCTGTGCTTTCAGATATCTGCTTTCATATGACCTTG 430

RESULT 15
LOCUS HUMHIS1X 480 bp mRNA linear PRI 31-DEC-1994
DEFINITION Human histatin 1 (HIS1) mRNA, complete cds.
ACCESSION M26664
VERSION M26664.1 GI:292143

KEYWORDS	histatin 1.
SOURCE	Homo sapiens parotid gland cDNA to mRNA
ORGANISM	Homo sapiens

SOURCE ORGANISM	CDNA TO mRNA
Homo sapiens	parotid gland
Homo sapiens	

ORGANISM	Homo sapiens
Pathogen	

REFERENCE
AUTHORS
1 (bases 1 to 480)
Sabbatini, M. and ...
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 480)

NO. WORKS	TITLE
3	SADDALHILL, L.M. and A. H. STATINS, a family

encoded by at least two loci (HIS1 and HIS2)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 03:13:40 ; Search time 341 seconds
(without alignments)
8790.058 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 1331
Sequence: 1 ctatgttttagatcaaacg.....gcctttccatgtctacttg 1331

Scoring table: IDENTITY_NMC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1330	99.9	1331	21	AAAA8963 Human lysine-rich
2	253	19.0	552	21	AAAA8964 Human statherin DN
3	209.6	15.7	2121	21	AAC76665 Human ORFX ORF2220
4	187.4	14.1	221	21	AAAA2335 Human secreted exp
5	86.6	6.5	390	22	AA191635 Human polynucleoti
6	86.2	6.5	857	22	AAH98658 Human EST-derived
7	78.8	5.9	491	21	AAH98658 Human basic histid
8	78.6	5.9	126	16	AAT24320 Human gene signatu
9	76.4	5.7	90	16	AAT23140 Human gene signatu

10	74.6	5.6	9881	24	ABL54353	Chemically treated
11	60.2	4.5	516	20	AAV89525	EST clone CP289.
12	56.4	4.2	60	24	ABN43273	Human spliced tran
13	55.8	4.2	9881	24	ABL54354	Chemically treated
14	48.4	3.6	6056	24	ABL33026	Human immune syste
15	47	3.5	61020	22	AA546787	Tumour suppressor
16	46.8	3.5	10279	24	ABL92276	Chemically treated
17	46.8	3.5	10279	24	ABL33590	Human immune syste
18	46.8	3.5	10279	24	ABD22327	Chemically treated
19	46.8	3.5	51345	22	AA161373	Human immune syste
20	46	3.5	6131	24	ABL32891	Soybean 318013 reg
21	44.8	3.4	545	22	AAH70126	Human immune syste
22	44.6	3.4	536	22	AAH91173	Human cervical can
23	44	3.3	16766	24	ABL34157	Human immune syste
24	43.8	3.3	5461	24	ABO66977	Human immune syste
25	43.8	3.3	5461	24	ABL32998	Human immune syste
26	43.8	3.3	5647	24	ABL70355	Human immune syste
27	43.8	3.3	5647	24	ABL33566	Chemically treated
28	43.8	3.3	5647	24	AA561320	Human immune syste
29	43.4	3.3	8577	24	ABL33787	Human immune syste
30	43.4	3.3	19659	24	ABL32766	Human immune syste
31	42.8	3.2	375	23	ABV44911	Human prostate exp
32	42.8	3.2	8946	24	ABL32910	Human immune syste
33	42.6	3.2	3680	22	AAH14565	Human immune syste
34	42.6	3.2	7823	22	AAH72614	Human CDNA sequenc
35	42.4	3.2	18183	22	AA546279	Human cervical can
36	42.4	3.2	18183	24	ABL70111	Tumour suppressor
37	42.4	3.2	18183	24	ABK31158	Chemically treated
38	42.2	3.2	1160	19	AAV44863	Signal transductio
39	42.2	3.2	1160	22	AAV98460	Clone CR201_1 codi
40	42.2	3.2	6055	22	AA536246	Human CDNA clone C
41	42.2	3.2	7809	22	AA536246	Human cardiovascular
42	42.2	3.2	7809	22	AA536245	Human cardiovascular
43	42	3.2	707	22	AA124056	Human breast cance
44	42	3.2	15373	24	ABL32466	Human immune syste
45	41.8	3.1	7171	24	ABN80216	Human chemically m

ALIGNMENTS

RESULT 1
ID AAA48963 standard: CDNA, 1331 BP.

AC AAA48963;

XX 06-OCT-2000 (first entry)

XX Human Lysine-rich statherin cDNA from Incyte clone 2820214.

XX Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;
XX precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;
XX asthma; allergy; diabetes mellitus; fungal; bacterial infection;
XX cancer; leukemia; adenocarcinoma; melanoma; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 871..1158 /tag- a

XX FT /product- Lysine-rich-statherin

XX PN WO200024779-A1.

XX PD 04-MAY-2000.

XX PF 22-OCT-1999; 99WO-US24046.

XX PR 23-OCT-1998; 98US-0155209.

XX PA (INCYTE-) INCYTE PHARM INC.

PI Tang YF, Corley NC, Guegler KJ, Patterson C;
XX
DR MPI: 2000-350699/30.
DR P-PSDB: AAY94526.

Purified polypeptide used for treating or preventing a disorder
characterized by expression or activity of lysine-rich statherin
proteins.

Claim 3; Page 70; 75pp; English.

The present sequence is human lysine-rich statherin protein (LRSP)
cDNA from Inyte clone 2820214. This sequence was identified
through analysis of a cDNA library of breast tumor tissue
(BRSTN014). The LRSP sequence was found to have homology with
human statherin (AAY94527) and human basic histidine-rich protein
(AAY94528). Human statherin is a phosphoprotein that acts as an
inhibitor of precipitation of calcium phosphate salts in the oral
cavity. The LRSP polypeptide and its antagonists may be useful for
treating or preventing disorders associated with the activity of
LRSP. Such disorders include autoimmune/inflammatory disorders (for
example AIDS, allergies, asthma, diabetes mellitus), bacterial and
fungal infection and cancers (such as leukemia, adenocarcinoma,
melanoma). Antibodies to LRSP may be useful for diagnosis of
the above disorders.

Sequence 1331 BP; 397 A; 258 C; 228 G; 447 T; 1 other;

Query Match 99.9%; Score 1330; DB 21; Length 1331;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTAATGTTTAAAGTAAAGATGACCGGTAGCTGTCTACATGTCACCAAGTAAATTT 60
DB 1 CTAATGTTTAAAGTAAAGATGACCGGTAGCTGTCTACATGTCACCAAGTAAATTT 60
QY 61 AAGAGTCTTACAAATGTCGATTTTGCATGCGCATAGGGAATGTAAGTTACTTGGC 120
DB 61 AAGAGTCTTACAAATGTCGATTTTGCATGCGCATAGGGAATGTAAGTTACTTGGC 120
QY 121 TGGAAATTTATCAGACTTGTAGTAAACAAGTTGAGAGTTTACACATGAGGGGAATATG 180
DB 121 TGGAAATTTATCAGACTTGTAGTAAACAAGTTGAGAGTTTACACATGAGGGGAATATG 180
QY 181 AGCCCCCTTAAGGCTTAACAAATAATCAGATCTGAGATAGTGGCTATGCGTCCCA 240
DB 181 AGCCCCCTTAAGGCTTAACAAATAATCAGATCTGAGATAGTGGCTATGCGTCCCA 240
QY 241 GGCCTAATTTGGGACACTTTTCTGATGCTTTGAGAGTACTTCTTTTGAAGAAA 300
DB 241 GGCCTAATTTGGGACACTTTTCTGATGCTTTGAGAGTACTTCTTTTGAAGAAA 300
QY 301 TTTTCATTTCTGCTGCTGATGCTATATCTCCCTTATAGAGCATTTGATTTCTTCC 360
DB 301 TTTTCATTTCTGCTGCTGATGCTATATCTCCCTTATAGAGCATTTGATTTCTTCC 360
QY 361 TTTTCTGGGAAATGCTCCATTTAGCATTTTGTGATGTCGACTATGCAATAT 420
DB 361 TTTTCTGGGAAATGCTCCATTTAGCATTTTGTGATGTCGACTATGCAATAT 420
QY 421 TGGTAATGCGCTTATGTTGTAATACAGATGTAATTAACGCTTACATAATCTACA 480
DB 421 TGGTAATGCGCTTATGTTGTAATACAGATGTAATTAACGCTTACATAATCTACA 480
QY 481 CTTGGATTTGCTGCACTCTACCAATAGCTTTTGAATGACTGAAGTGTAAACAGAA 540
DB 481 CTTGGATTTGCTGCACTCTACCAATAGCTTTTGAATGACTGAAGTGTAAACAGAA 540
QY 541 AGAGGCAATGCTGCAAGAGATAGTAAATATTTTGTACTTTATCTGAATCCAAG 600
DB 541 AGAGGCAATGCTGCAAGAGATAGTAAATATTTTGTACTTTATCTGAATCCAAG 600
QY 601 ATGCTGCTCCCTGACAGTGTGTTTCTTACAGATCTCATGTAATCCCTCTGGGA 660
DB 601 ATGCTGCTCCCTGACAGTGTGTTTCTTACAGATCTCATGTAATCCCTCTGGGA 660

DB 601 ATGCTGCTCCCTGACAGTGTGTTTCTTACAGATCTCATGTAATCCCTCTGGGA 660
QY 661 GCACAGGACAGTATAGTAACTCTCCATTTCTTTGTTTGTATTAAGACAGACTCT 720
DB 661 GCACAGGACAGTATAGTAACTCTCCATTTCTTTGTTTGTATTAAGACAGACTCT 720
QY 721 GTCTCAAAAAAAGACATTTATATTAACATCTTATTAAGAGCCCTTAATTTCTATC 780
DB 721 GTCTCAAAAAAAGACATTTATATTAACATCTTATTAAGAGCCCTTAATTTCTATC 780
QY 781 TGAAGGACACTGTTTTTTTTTAAACAGTTAAGTACGATGTCACAGCAAAATTTCT 840
DB 781 TGAAGGACACTGTTTTTTTTTAAACAGTTAAGTACGATGTCACAGCAAAATTTCT 840
QY 841 GATCAGATAGTCCCTGTCACAGTACCAATGTTGTTTCATTAAGTGGGAAGAAACAG 900
DB 841 GATCAGATAGTCCCTGTCACAGTACCAATGTTGTTTCATTAAGTGGGAAGAAACAG 900
QY 901 CATTTTAAAGTACTTTTGGGAGACTGATTTGAGTAATTAATTAACCTGCTCCCTT 960
DB 901 CATTTTAAAGTACTTTTGGGAGACTGATTTGAGTAATTAATTAACCTGCTCCCTT 960
QY 961 AAGAAAAAACCCTTCCACCTTACTGTGCTATTAATCCCTTACTTCCAAAGTTA 1020
DB 961 AAGAAAAAACCCTTCCACCTTACTGTGCTATTAATCCCTTACTTCCAAAGTTA 1020
QY 1021 ATTATCTTATTTGATGATTTGCTTTTATACCAAGAGCTTATCAGCAGTTCCAGAAC 1080
DB 1021 ATTATCTTATTTGATGATTTGCTTTTATACCAAGAGCTTATCAGCAGTTCCAGAAC 1080
QY 1081 AACCACTATACGACACACATACCAACCAATATACCAATATAGCTTAAATATCATC 1140
DB 1081 AACCACTATACGACACACATACCAACCAATATACCAATATAGCTTAAATATCATC 1140
QY 1141 AGTAAGTGAAGGACATGATTTTATGAGCTTATGAGCAATATGCACTTCTACATCATAT 1200
DB 1141 AGTAAGTGAAGGACATGATTTTATGAGCTTATGAGCAATATGCACTTCTACATCATAT 1200
QY 1201 TCTCATCTTTCATATACCATATACACATACATACATTTTGTAGATCATTAAGAGCAAT 1260
DB 1201 TCTCATCTTTCATATACCATATACACATACATACATTTTGTAGATCATTAAGAGCAAT 1260
QY 1261 GCGAATGTAAACCTTATATTTACTGATGATCTTTGTTCCAGATACTTGCCTTTTCC 1320
DB 1261 GCGAATGTAAACCTTATATTTACTGATGATCTTTGTTCCAGATACTTGCCTTTTCC 1320
QY 1321 AATGTCACCTG 1331
DB 1321 AATGTCACCTG 1331

RESULT 2
AAA48964
ID AAA48964 standard; DNA; 552 Bp.
XX
AC AAA48964;
AC
DT 06-OCT-2000 (first entry)
XX
DE Human statherin DNA.
XX
KW Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;
KW precipitation inhibitor; autoimmune; inflammatory disorder; AIDS;
KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;
KW cancer; leukemia; adenocarcinoma; melanoma; ds.
XX
OS Homo sapiens.
XX
FH Key 73..261 Location/Qualifiers
FT CDS
FT
FT /product= Statherin

xx WO200024779-A1.
 xx 04-MAY-2000.
 xx 22-OCT-1999; 99WO-0524046.
 xx 23-OCT-1998; 98US-0155209.
 xx (INCY-) INCYTE PHARM INC.
 xx Tang YT, Corley NC, Guejler KJ, Patterson C;
 xx WPI: 2000-350699/30.
 xx P-PSDB: AAY94527.
 xx Purified polypeptide used for treating or preventing a disorder
 xx characterized by expression or activity of lysine-rich statherin
 xx proteins.
 xx Disclosure; Page 71; 75pp; English.
 xx The present invention relates to human lysine-rich statherin protein
 xx (LRSP)(AAY94526). The cDNA sequence encoding this protein was identified
 xx through analysis of a cDNA library of breast tumour tissue (BRSTM014).
 xx The LRSP sequence was found to have homology with the DNA of human
 xx statherin (the present sequence) and human basic histidine-rich protein
 xx (AAY94528). Human statherin is a phosphoprotein that acts as an
 xx inhibitor of precipitation of calcium phosphate salts in the oral cavity.
 xx The LRSP polypeptide and its antagonists may be useful for treating or
 xx preventing disorders associated with the activity of LRSP. Such
 xx disorders include autoimmune/inflammatory disorders (for example AIDS,
 xx allergies, asthma, diabetes mellitus), bacterial and fungal infection
 xx and cancers (such as leukemia, adenocarcinoma, melanoma). Antibodies to
 xx LRSP may be useful for diagnosis of the above disorders.
 xx Sequence 552 BP; 172 A; 122 C; 78 G; 180 T; 0 other;
 xx
 xx Query Match 19.0%; Score 253; DB 21; Length 552;
 xx Best Local Similarity 88.4%; Pred. No. 2e-53;
 xx Matches 274; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
 xx
 xx 1022 TTATCTTATTTGGATGATTTGCTTTATACCAAGAGCTTATCAGGCTCCAGAACCA 1081
 xx 150 TTGGCGAGATTTGAGATTCGGTTATGGGTATAGCCCTTATCAGCAGATTCAGAACCA 209
 xx 1082 ACCACTATAGGCAACCAATACCAACCAACCAATATACCAATATATATATATATCA 1141
 xx 210 ACCACTATATACCAACCAATACCAACCAACCAATATATATATATATATATATCA 269
 xx 1142 GTAAGTGCAGAGATGATTTATGAGGCTGATTTGCGAAATAGCACTTCTATCCATATAT 1201
 xx 270 GTAAGTGCAGAGATGATTTATGAGGCTGATTTGCGAAATAGCACTTCTATCCATATAT 329
 xx 1202 CTGATCTTATATACCAAT 1261
 xx 330 CTGATCTTATATACCAAT 389
 xx 1262 CGAATGTAAACCACTTAT 1321
 xx 390 CAAATGAAACCACTTAT 449
 xx 1322 ATGTCACTTG 1331
 xx 450 TTGTCACTTG 459
 xx
 xx RESULT 3
 xx AAC76665
 xx ID AAC76665 standard; cDNA: 2121 BP.
 xx AC AAC76665;
 xx

DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF2220 polynucleotide sequence SEQ ID NO:4439.
 xx Human: open reading frame; ORFX; detection; cytosolic; hepatotropic;
 xx vulnery; antiparietal; antiparkinsonian; noctropic; neuroprotective;
 xx anticonvulsant; osteoplastic; antidiabetic; immunosuppressant; cardiac;
 xx immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 xx hypotensive; dermatological; immunosuppressive; antidiabetic;
 xx antiviral; antibacterial; antifungal; antineumatic; antihypertensive;
 xx antineumatic; gene therapy; cancer; proliferative disorder; hypertension;
 xx neurodegenerative disorder; osteoarthritis; graft vs host disease;
 xx cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 xx cholesterol ester storage; systemic lupus erythematosus; infection;
 xx severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 xx allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 xx bone damage; cartilage damage; nocturnal haemoglobinuria; burn; wound;
 xx thrombosis; contraceptive; ss.
 xx
 xx Homo sapiens.
 xx
 xx WO200058473-A2.
 xx 05-OCT-2000.
 xx 31-MAR-2000; 2000WO-US08621.
 xx 31-MAR-1999; 99US-0127607.
 xx 02-APR-1999; 99US-0127636.
 xx 05-APR-1999; 99US-0127728.
 xx 30-MAR-2000; 2000US-0540763.
 xx (CURA-) CURAGEN CORP.
 xx
 xx Shinkets RA, Leach M;
 xx WPI: 2000-602363/57.
 xx P-PSDB: AAB42456.
 xx Novel nucleic acids and peptides derived from open reading frame X,
 xx useful for treating e.g. cancers, proliferative disorders,
 xx neurodegenerative disorders and cardiovascular disease.
 xx
 xx Claim 5; Page 3631-3632; 5507pp; English.
 xx
 xx AAC74446 encode the proteins given in AAB40237 to AAB43397,
 xx which represent the human ORFX open reading frames 1 to 3161. The ORFX
 xx sequences have activities such as: cytostatic; hepatotropic; vulnery;
 xx antiparietal; antiparkinsonian; noctropic; neuroprotective;
 xx osteoplastic; anticonvulsant; antidiabetic; immunosuppressant;
 xx immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
 xx antidiabetic; hypotensive; dermatological; immunosuppressive;
 xx antineumatic; antihypertensive; antidiabetic; antidiabetic;
 xx antihypertensive; antineumatic. The sequences can be used for determining
 xx the presence of or predisposition to, or preventing or treating
 xx pathological conditions associated with an ORFX-associated disorder. The
 xx nucleic acids can be used to express ORFX proteins in gene therapy
 xx vectors. The proteins and nucleic acids may be used to treat cancers,
 xx proliferative disorders, neurodegenerative disorders, osteoarthritis,
 xx graft vs host disease, cardiovascular disease, diabetes mellitus,
 xx erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 xx bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 xx allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 xx nocturnal haemoglobinuria, antineumatic disease; to enhance
 xx coagulation; to inhibit thrombosis; and as a contraceptive.
 xx
 xx Sequence 2121 BP; 617 A; 430 C; 341 G; 732 T; 1 other;
 xx
 xx Query Match 15.7%; Score 209.6; DB 21; Length 2121;
 xx Best Local Similarity 81.5%; Pred. No. 2.2e-42;
 xx Matches 255; Conservative 0; Mismatches 55; Indels 3; Gaps 1;
 xx

PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI: 2001-514838/56.
 DR P-PSDB: AA011704.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 1: SEQ ID NO 11695; 13999p + Sequence listing; English.
 XX
 CC The invention relates to human polynucleotides (AA17994-AA19341) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SO Sequence 390 BP: 127 A; 77 C; 79 G; 106 T; 1 other:
 XX
 Query Match 6.5%; Score 86.6; DB 22; Length 390;
 Best Local Similarity 75.8%; Pred. No. 6.7e-12;
 Matches 122; Conservative 0; Mismatches 34; Indels 5; Gaps 1;
 XX
 OY 812 AGTACGATGTCACAGACAAATTTTCATCAGATAGTCCCTGCAACAGTACCAAA 871
 DB 152 AGAAAAATGAAATTTTCCAAATATTTCCGATCAGAGAAATC-----ACAAGACGACAAA 206
 OY 872 TGTGCTTCAATAGTGGGAAAGAAACAGCATTTTAAAGTAATTTTGGGAGACTGAT 931
 DB 207 TGTGCTTCAATAGTGGGAAAGAAACAGCATTTTAAAGTAATTTTGGGAGACTGAT 931
 OY 932 TGAGTAAATATTAATTAATCTGCTGCTTCCCTTAAGAAAAAAA 972
 DB 267 TGAGTAAATATTAATTAATCTGCTGCTTCCCTTAAGAAAAAAA 307
 XX
 RESULT 6
 AAH98658
 ID AAH98658 standard; cDNA; 857 BP.
 XX
 AC AAH98658;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Human EST-derived coding sequence SEQ ID NO: 515.
 XX
 KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200154477-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 25-JAN-2001; 2001WO-US02687.
 XX
 PR 25-JAN-2000; 2000US-0491404.
 PR 17-JUL-2000; 2000US-0617746.
 PR 03-AUG-2000; 2000US-0631451.

PR 15-SEP-2000; 2000US-0663870.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
 PI Cao Y, Drmanac RA, Zhang J, Werhman T;
 XX
 DR WPI: 2001-476164/51.
 DR P-PSDB: AAM23999.
 XX
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -
 XX
 PS Claim 1: Page 537; 1275pp; English.
 XX
 CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.
 CC
 XX
 SO Sequence 857 BP: 231 A; 194 C; 169 G; 263 T; 0 other:
 XX
 Query Match 6.5%; Score 86.2; DB 22; Length 857;
 Best Local Similarity 62.5%; Pred. No. 1.1e-11;
 Matches 150; Conservative 0; Mismatches 89; Indels 1; Gaps 1;
 XX
 OY 1093 CACACCATACACACACCAATACCAATATAGCTTTAAATATCAGTACGATCAGG 1152
 DB 510 CACATGAGGCTATATATCAATTAATCTGATGACAAATGATATCTAGTATCATG 569
 OY 1153 ACATGATATGAGGCTGATGAGCAATATACGACTTCTCATCATATCTCTTCA 1212
 DB 570 GCATGATTTGAGAGGTTTACGAGCAATTCCTTTGGACTCGTGTATCTCATTTG 629
 OY 1213 TACCATATACACTTACTTACTTCTTTGATGATCATCTAAGAGCAATGC-GAATGTA 1271
 DB 630 TACCGCATACACTACCACTGCTTTTGAAGATTAATCAATAGGCAATGCAATTAAG 689
 OY 1272 ACCCATATATTTACGATACCTTTGTTGTTCCAGATACCTGCTTTTCAATGCTCTG 1331
 DB 690 AATACCATGATTTAGTGAATTCGTGTTTCAGATACCTTCCCTTCTTAATATCATTTG 749
 XX
 RESULT 7
 AAA48965
 ID AAA48965 standard; DNA; 491 BP.
 XX
 AC AAA48965;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human basic histidine-rich protein DNA.
 XX
 KW Lysine-rich statherin protein; LRSP; acidic phosphoprotein; human;
 KW precipitation inhibitor; autoimmunity; inflammatory disorder; AIDS;
 KW asthma; allergy; diabetes mellitus; fungal; bacterial infection;
 KW cancer; leukemia; adenocarcinoma; melanoma; ds.
 XX
 OS Homo sapiens.
 XX
 PN Key
 XX
 FT CDS Location/Qualifiers
 FT 38..193
 FT /*tag= a
 FT /product= Basic_histidine-rich_protein
 XX
 PD WO200024779-A1.
 XX
 PD 04-MAY-2000.

PF 22-OCT-1999; 99WO-US24046.
 XX
 PR 23-OCT-1998; 98US-0155209.
 XX
 PA (INCYTE) INCYTE PHARM INC.
 XX
 PI Tang YT, Corley NC, Guegler KJ, Patterson C;
 XX
 DR WPI: 2000-350699/30.
 XX
 P-PSDB: AAY94528.
 XX
 PT Purified polypeptide used for treating or preventing a disorder
 PT characterized by expression or activity of lysine-rich statherin
 PT proteins -
 XX
 PS Disclosure; Page 72; 75pp; English.
 XX
 CC The present invention relates to human lysine-rich statherin protein
 CC (LRSP) (AAY94528). The cDNA sequence encoding this protein was identified
 CC through analysis of a cDNA library of breast tumour tissue (BRSTM014).
 CC The LRSP sequence was found to have homology with human statherin
 CC protein (AAY94527) and human basic histidine-rich protein (the present
 CC sequence). Human statherin is a phosphoprotein that acts as an inhibitor
 CC of precipitation of calcium phosphate salts in the oral cavity. The LRSP
 CC polypeptide and its antagonists may be useful for treating or preventing
 CC disorders associated with the activity of LRSP. Such disorders include
 CC autoimmune/inflammatory disorders (for example AIDS, allergies, asthma,
 CC diabetes mellitus), bacterial and fungal infection and cancers (such as
 CC leukemia, adenocarcinoma, melanoma). Antibodies to LRSP may be useful
 CC for diagnosis of the above disorders.
 XX
 SQ Sequence 491 BP; 151 A; 90 C; 87 G; 163 T; 0 other;
 XX
 Query Match 5.9%; Score 78.8; DB 21; Length 491;
 Best Local Similarity 60.6%; Pred. No. 6.5e-10;
 Matches 146; Conservative 0; Mismatches 93; Indels 2; Gaps 1;
 XX
 QY 1093 CACACACATACCAACCAATACCAATATACGTTTAAATATCATCATGACGAGG 1152
 DB 153 CACATCGAGCGCTATGATCAATATATCTGTATGACATGATATCTCTGATATCATGAG 212
 XX
 QY 1153 ACATGATTTATGAGGCTTATGGAATATGACATTCATCATCATGATCTTCA 1212
 DB 213 GCATGATTTATGAGGCTTATGGAATATGACATTCATCATCATGATCTTCA 272
 XX
 QY 1213 TACCATATCACACTACTACCTTTTTCGTTTGAAGCAATATCATGAGCAATGATAA 1270
 DB 273 TACCGCATCACACTACTACTGCTTTTTCGTTTGAAGCAATATCATGAGCAATGATAA 332
 XX
 QY 1271 AACCCCTATATATTTACTGATATCTTGTGTTCCAGATATCTGCTTTTCAATGCACTT 1330
 DB 333 GAATATACCATGATTAAGTATGATGCTGTTTCAGATATCTGCTTTTCAATGCACTT 392
 XX
 QY 1331 G 1331
 DB 393 G 393
 XX
 RESULT 8
 AAT24320
 ID AAT24320 standard; cDNA to mRNA; 126 BP.
 XX
 AC AAT24320;
 XX
 DE 22-SEP-1996 (first entry).
 XX
 DE Human gene signature HUMGS06347.
 XX
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 XX
 OS Homo sapiens.

XX
 PN WO9514772-A1.
 XX
 PD 01-JUN-1995.
 XX
 PF 11-NOV-1994; 94WO-JP01916.
 XX
 PR 12-NOV-1993; 93JP-0355504.
 XX
 PA (MATSUBARA) MATSUBARA K.
 XX
 PI (OKUBO) OKUBO K.
 XX
 PI Matsubara K, Okubo K;
 XX
 DR WPI: 1995-206931/27.
 XX
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 XX
 PS Claim 1; Page 1582; 2245pp; Japanese.
 XX
 CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-726837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 XX
 SQ Sequence 126 BP; 44 A; 16 C; 25 G; 34 T; 7 other;
 XX
 Query Match 5.9%; Score 78.6; DB 16; Length 126;
 Best Local Similarity 86.6%; Pred. No. 4.6e-10;
 Matches 84; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 XX
 QY 859 CAAAGTAGCAATATGCTTTCATTAAGTGGGAAGCAACATTTTAAGTAACTTTT 918
 DB 14 CAAGACGAGCAATATGCTTTCATTAAGTGGGAAGCAACATTTTAAGTAACTTTT 73
 XX
 QY 919 TGGGACAGCTGATTTGAGTAATATAAAGCTGCTCT 955
 DB 74 TGGGACAGCTGATTTGAGTAATATAAAGCTGCTCT 110
 XX
 RESULT 9
 AAT23140
 ID AAT23140 standard; cDNA to mRNA; 90 BP.
 XX
 AC AAT23140;
 XX
 DE 23-AUG-1996 (first entry)
 XX
 DE Human gene signature HUMGS04875.
 XX
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9514772-A1.
 XX
 PD 01-JUN-1995.


```

XX PF 11-NOV-1994; 94WO-JP01916.
XX XX
XX PR 12-NOV-1993; 93JP-0355504.
XX XX
XX PA (MATS/) MATSUBARA K.
XX PA (OKUB/) OKUBO K.
XX XX
XX PI Matsubara K, Okubo K;
XX DR
XX WPI; 1995-206931/27.
XX
XX PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
XX PT for diagnosis of abnormal cell function, by preparing cDNA that
XX PT reflects relative abundance of corresp. mRNA in specific human
XX PT tissues
XX
XX Claim 1; Page 1297; 2245pp; Japanese.
XX
XX A single-stranded DNA (or its complementary strand or the corresp.
XX CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
XX CC given in AA119001-T26837 and which is able to hybridise to part of
XX CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
XX CC sequences were obtained from 3'-directed cDNA libraries prepared
XX CC from various human tissues; synthesis of cDNA was initiated from the
XX CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
XX CC untranslated sequence is unique to a particular mRNA species, almost
XX CC all the 3'-oriented cDNAs hybridise with specific mRNA species, each
XX CC is constructed so as to reflect accurately the relative abundance of
XX CC different mRNAs in the particular tissue from which it was derived.
XX CC The appearance frequency of a given GS in a cDNA library can be
XX CC determined (esp. using primers and probes derived from the GS
XX CC sequences) as a means of diagnosing abnormal cell function or for
XX CC recognising different cell types.
XX
XX Sequence 90 BP; 30 A; 12 C; 17 G; 31 T; 0 other;
XX
XX Query Match 5.7%; Score 76.4; DB 16; Length 90;
XX Best Local Similarity 97.8%; Pred. No. 1.4e-09;
XX Matches 88; Conservative 0; Mismatches 1; Indels 1; Gaps 1
XX
XX QY 392 GATCTTTGATGTCGACCTAATGCCATTATGTGTAATGCC-GTTATTGGTAATACAGCAT 450
XX DB 1 GATCTTTGATGTCGACCTAATGCCATTATGTGTAATGCCGTTATTGGTAATACAGCAT 60
XX
XX QY 451 AGTTAATAAATCTGTACAGTAATCTACA 480
XX DB 61 AGTTAATAAATCTGTACAGTAATCTACA 90
XX
XX RESULT 10
XX ID ABL54353 standard; DNA; 9881 BP.
XX AC ABL54353;
XX
XX DT 29-JUL-2002 (first entry)
XX
XX DE Chemically treated apoptosis gene #27.
XX
XX KW Apoptosis; HIV; Bloom syndrome; cardiopathy;
XX KW neurodegenerative disorder; Herpes simplex virus; renal ischaemia;
XX KW amyotrophic lateral sclerosis; cancer; ds.
XX
XX OS Unidentified.
XX
XX MO200177164-A2.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-EP03969.
XX
XX 06-APR-2000; 2000DE-1019058.
XX

```

```

PR 07-APR-2000: 2000DE-1019173.
PR 30-JUN-2000: 2000DE-1032529.
PR 01-SEP-2000: 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-017444/02.
XX
XX Chemically modified sequences of genes associated with apoptosis are
PT useful to determine methylation patterns of genomic DNA samples for
PT diagnosis of associated diseases such as cancer
XX
XX Claim 1; Seq ID #53; 24bp; English.
PS
CC This invention relates to chemically pre-treated DNA of genes
CC associated with apoptosis. The nucleic acids are used to allocate
CC patients for specific therapy for HIV infection, Bloom syndrome,
CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus
CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours
CC and cancers. This nucleotide sequence represents a chemically
CC treated apoptosis gene. Even SEQ ID numbers are the complementary
CC DNA strands to the cod SEQ ID numbers. The sequence data for this
CC patent is not represented in the printed specification but is based on
CC information supplied by the European patent office.
XX
XX Sequence 9881 BP; 2606 A; 196 C; 2308 G; 4771 T; 0 other;
SQ
Query Match 5.6%; Score 74.6; DB 24; Length 9881;
Best Local Similarity 78.8%; Pred. No. 2e-08;
Matches 89; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 860 AACAGTAGCAAAATGCTTTCATTAAGCGGAGAAACAGCATTTTAAAGTACTTTT 919
DB 5771 AAGAGTAGTAATAATGCTTTATTATTAAGCGGAGAAAGTACTATTTAATTAATTTT 5830
QY 920 GGGAGACGTGATTGGATTAATAATAAACTCGGCTCCCTTAAGAAAAAAA 972
DB 5831 GGGAGATTGAATTGAGTAATAATAATAAAATTTTTCGTTAATAATAAATAA 5883
RESULT 11
AAVB9525
ID AAVB9525 standard; cDNA; 516 BP.
AC AAVB9525;
XX
XX 15-FEB-1999 (first entry)
DT
XX
XX EST clone CP289.
DE
XX
XX Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; actinin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
XX
XX Homo sapiens.
OS
XX
XX WO9845436-A2.
PN
XX
XX 15-OCT-1998.
PD
XX
XX 10-APR-1998; 98WO-US06955.
PF
XX 10-APR-1997; 97US-0838821.
PR
XX
XX (GEMV ) GENETICS INST INC.
PA
XX
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
XX WPI; 1999-070077/06.
XX

```

XX New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
XX ovary, pituitary, retina and colon cDNA libraries.

PS Claim 1: Page 245; 618pp; English.

CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cachectin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.

XX Sequence 516 BP; 156 A; 181 C; 97 G; 82 T; 0 other:

Query Match 4.5%: Score 60.2; DB 20; Length 516;
Best Local Similarity 89.0%: Pred. No. 2.9e-05;
Matches 65; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1047 TATACCAAGAGCGCTTATCAGCCAGTTCAGAACACCACTATACGACCAATACCA 1106
DB 139 TATGGCGATATGCGCTTATCAGCCAGTTCAGAACACCACTATACGACCAATACCA 198
OY 1107 CGACATATACCAAC 1119
DB 199 CCACAAATACCAAC 211

RESULT 12
ID ABN43273 standard; DNA: 60 BP.

XX ABN43273:
AC 15-JUL-2002 (first entry)

XX Human spliced transcript detection oligonucleotide SEQ ID NO:16021.
DE Human spliced transcript detection oligonucleotide detection: RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.

XX Homo sapiens.
OS
XX MO200210449-A2.
XX
XX 07-FEB-2002.

XX 20-JUL-2001; 2001WO-1B01903.
XX
XX 28-JUL-2000; 2000US-221607P.
XX 02-MAY-2001; 2001US-287724P.

XX (COMP-) COMPUGEN INC.
XX
XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
XX WPI; 2002-257383/30.

XX New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes

XX Example 1: SEQ ID 16021; 47pp; English.

XX The present invention describes oligonucleotide libraries for detecting

CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridizing selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biologically sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialized mini-
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 60 BP; 22 A; 25 C; 4 G; 9 T; 0 other:
Query Match 4.2%: Score 56.4; DB 24; Length 60;
Best Local Similarity 98.3%: Pred. No. 0.00012;
Matches 57; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1059 CCTATACGACGCTTCCAGAACACCACTATACGACCAATACCAATACCA 1116
DB 3 CCTATACGACGCTTCCAGAACACCACTATACGACCAATACCAATACCAATACCA 60

RESULT 13
ID ABL54354/C standard; DNA: 9881 BP.

XX ABL54354:
AC 29-JUL-2002 (first entry)

XX Chemically treated apoptosis gene complementary to gene #27.
DE
XX
XX Apoptosis; HIV; Bloom syndrome; cardiopathy;
XX neurodegenerative disorder; Herpes simplex virus; renal ischaemia;
XX amyotrophic lateral sclerosis; cancer; ds.

XX Unidentified.
OS
XX MO200177164-A2.
XX
XX 18-OCT-2001.

XX 06-APR-2001; 2001WO-EP03969.
XX
XX 06-APR-2000; 2000DE-1019058.
XX 07-APR-2000; 2000DE-1019173.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-017444/02.

XX Chemically modified sequences of genes associated with apoptosis are
PT useful to determine methylation patterns of genomic DNA samples for
PT diagnosis of associated diseases such as cancer

XX Claim 1: Seq ID #54; 24pp; English.

XX This invention relates to chemically pre-treated DNA of genes
 CC associated with apoptosis. The nucleic acids are used to allocate
 CC patients for specific therapy for HIV infection, Bloom syndrome,
 CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus
 CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours
 CC and cancers. This nucleotide sequence represents a chemically
 CC treated apoptosis gene. Even SEQ ID numbers are the complementary
 CC DNA strands to the odd SEQ ID numbers. The sequence data for this
 CC patent is not represented by the printed specification but is based on
 CC information supplied by the European patent office.

XX Sequence 9881 BP; 2874 A; 196 C; 2093 G; 4718 T; 0 other;

Query Match 4.2%; Score 55.8; DB 24; Length 9881;
 Best Local Similarity 67.1%; Pred. No. 0.001;

Matches 96; Conservative 0; Mismatches 42; Indels 5; Gaps 1;

OY 830 CAATATTTTCGATCAGATAGTCCCTGTCACAGTCAATGTTCTTAAGTGG 889
 DB 4136 CAATATTTTCGATCAGATAGTCCCTGTCACAGTCAATGTTCTTAAGTGG 4082
 OY 890 GAAGAAAACAGCATTTTAAAGTACTTTTGGAGAGCTGATTGATATATAAACTC 949
 DB 4081 AAAAAAACAACATTTTAAATTAATTTTAAAAAATAAATAATATAAACTT 4022
 OY 950 TGTCTCCCTTAAGAAAAA 972
 DB 4021 CAATCTTTCGCTAATATATATA 3999

RESULT 14

ABL33026
 ID ABL33026 standard; DNA: 6056 BP.

XX ABL33026;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 999.

XX Human; immune system disease; cytosine methylation; antileukemic;
 KM antiarteriosclerotic; antileukemic; cytosine; nucleotide;
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KM antineoplastic; antileukemic; antidiabetic; antiparasitic;
 KM antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KM gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX MPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 PS claim 1, SEQ ID NO 999; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 6056 BP; 1780 A; 54 C; 1079 G; 3143 T; 0 other;

Query Match 3.6%; Score 48.4; DB 24; Length 6056;
 Best Local Similarity 45.9%; Pred. No. 0.06;

Matches 166; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

OY 688 TTCTCTTTTGGTTTAAAGACAGACCTGCTCAAAAAAGACATTTATCAT 747
 DB 3741 TTTATGATTTTGGAAATGAAATTTTATTTTGTATTAATAATTAATAA 3800
 OY 748 ATACATCTTTATAGAGCCCTTAATTTCTATGAGGACGCTTTTAAACA 807
 DB 3801 AGTAGTGTATTTATAGATTTTGTATTTTGAATTTTGTATTAATAA 3860
 OY 808 GTTAGTACGATCTCAACAGCAATATTTCTGATCAGATAGTCCCTGCAACAGTAG 867
 DB 3861 ATTTTGTATTTGCTAAATATTAATAATTAATAATTTTGTATTAATAA 3920
 OY 868 CAATGTGTTCTCAATAGTGGAAAGAAACAGCATTTTAAAGTACTTTTGGAGACT 927
 DB 3921 GTATATATAGTTTATGATATTTTAAAGAAAAATGTTAAAGAAATTTGTTATTAAT 3980
 OY 928 GATTGTGATTAATAAAGCTGCTCCCTTAAGAAAAAACCTCCACCTTTAC 987
 DB 3981 TATTTTAAAGATTTGATATTTTATTTTGTATTTTAAAGCTTTATTTAAGATTAAT 4040
 OY 988 TGTCTCATTTATATCCCTTAGTTCCAAAGTTAATTTCTTATTTTGGATATTTGCTTTT 1047
 DB 4041 TTTTATATGTTGAATATATAGTTTAAATATGATTTTAAATGTTATTTAATAT 4100
 OY 1048 AT 1049
 DB 4101 TT 4102

RESULT 15

AAS46787/C
 ID AAS46787 standard; DNA: 61020 BP.

XX AAS46787;

XX 18-DEC-2001 (first entry)

XX Tumour suppressor gene derived chemically modified sequence #513.

XX Human; tumour suppressor gene; oncogene; antitumour; cytosine;
 KM cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KM cytosine methylation; ds.

XX Homo sapiens.

XX WO200168912-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02955.

XX 15-MAR-2000; 2000DE-1013847.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 14:15:20 ; Search time 92 Seconds
(without alignments)
4436.813 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 1331

Sequence: 1 ctatgttttagaatacaag.....gcctttccatgactctg 1331

Scoring table:

IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 88724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq.*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq.*
5: /cgn2_6/ptodata/1/lna/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/1/lna/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39.6	3.0	7218	1 US-08-232-463-14	Sequence 14, Appl
2	38.6	2.9	6265	4 US-09-129-112-3	Sequence 3, Appl
3	38	2.9	588	4 US-09-385-982-133	Sequence 133, App
4	37.2	2.8	5852	1 US-07-867-106-2	Sequence 2, Appl
5	37.2	2.8	246240	2 US-08-724-394A-20	Sequence 20, Appl
6	37.2	2.8	246240	2 US-08-724-394A-21	Sequence 21, Appl
7	37.2	2.8	246240	2 US-08-724-394A-22	Sequence 22, Appl
8	36.4	2.7	12730	4 US-09-004-838-91	Sequence 91, Appl
9	36.4	2.7	12730	4 US-09-004-838-124	Sequence 124, App
10	36.2	2.7	762	4 US-09-134-001C-1748	Sequence 748, App
11	36.2	2.7	8302	4 US-09-234-827B-1	Sequence 1, Appl
12	36	2.7	3718	4 US-09-424-283-6	Sequence 6, Appl
13	35.8	2.7	72928	3 US-09-009-913-1	Sequence 1, Appl
14	35.6	2.7	3780	4 US-09-134-001C-920	Sequence 920, App
15	35.6	2.7	19124	2 US-08-487-826B-13	Sequence 14, Appl
16	35	2.6	1474	4 US-08-821-994-64	Sequence 64, Appl
17	35	2.6	2192	4 US-08-714-918-14	Sequence 14, Appl
18	35	2.6	2192	4 US-09-265-315-14	Sequence 14, Appl
19	35	2.6	2192	4 US-09-265-315-14	Sequence 14, Appl
20	35	2.6	2192	4 US-09-265-315-14	Sequence 14, Appl
21	35	2.6	9048	3 US-08-973-273-4	Sequence 4, Appl
22	35	2.6	36651	4 US-09-738-894A-3	Sequence 3, Appl
23	34.8	2.6	4517	4 US-09-140-804-9	Sequence 9, Appl
24	34.8	2.6	11056	4 US-09-004-838-23	Sequence 23, Appl
25	34.8	2.6	15062	4 US-09-004-838-89	Sequence 89, Appl
26	34.8	2.6	62804	4 US-09-800-960-3	Sequence 3, Appl
27	34.4	2.6	2173	6 5168051-10	Patent No. 5168051

28	34.2	2.6	688	4 US-08-998-416-915	Sequence 915, App
29	34.2	2.6	1394	4 US-09-247-155-76	Sequence 76, Appl
30	34.2	2.6	1408	3 US-08-889-841B-22	Sequence 22, Appl
31	34.2	2.6	7228	2 US-08-850-049-128	Sequence 128, App
32	34.2	2.6	7228	2 US-08-850-049-129	Sequence 129, App
33	34.2	2.6	7228	2 US-08-050-478-128	Sequence 128, App
34	34.2	2.6	7228	2 US-08-050-478-129	Sequence 129, App
35	34.2	2.6	7228	4 US-09-414-117-128	Sequence 128, App
36	34.2	2.6	7228	4 US-09-414-117-129	Sequence 129, App
37	34.2	2.6	7228	4 US-09-678-437-128	Sequence 128, App
38	34.2	2.6	7228	4 US-09-678-437-129	Sequence 129, App
39	34.2	2.6	15581	3 US-08-646-538-35	Sequence 35, Appl
40	34.2	2.6	15581	4 US-09-503-222-35	Sequence 35, Appl
41	34	2.6	1744	4 US-09-511-625B-19	Sequence 19, Appl
42	34	2.6	17425	4 US-09-511-625B-5	Sequence 5, Appl
43	34	2.6	32768	4 US-08-961-527-71	Sequence 71, Appl
44	33.8	2.5	1786	1 US-07-920-430-19	Sequence 19, Appl
45	33.8	2.5	1786	1 US-08-066-299-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER-READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14
Query Match 3.0%, Score 39.6, DB 1, Length 7218;

Best Local Similarity 7.0%; Pred. No. 0.54;
Matches 18; Conservative 138; Mismatches 102; Indels 0; Gaps 0;

QY 545 GCATGCTCTCCAGAAAGATACCTATATTTTGGTACTTATCGAATCCAGATGC 604
DB 1037 GCTTGCTCGACGAGCGAGCTGCATGATTTTTTTTTTTTTTTTTTTTTT 1096
QY 605 TCGTCCCTCGAGGTGTTTCTTCTTACATCCTCTAATCCCTCGGAGCAC 664
DB 1097 TT 1156
QY 665 AGCAGATGATGACATCTCCATTTCTTGTGTTTAAAGACAGACGCTGCT 724
DB 1157 TT 1216
QY 725 CAAAAAAGACATTTATCATTAATACCTTATAGACCCCTAATTTCTTATCTGAA 784
DB 1217 TT 1276
QY 785 GGCACCTGTTTTTTTTTTT 802
DB 1277 TTTTTTTTTTTTTTTTTT 1294

RESULT 2

US-09-129-112-3/C
Sequence 3, Application US/09129112

Patent No. 6465716

GENERAL INFORMATION:

APPLICANT: Etzler, Marilyn E.

APPLICANT: Murphy, Judith B.

TITLE OF INVENTION: The Regents of the University of California

FILE REFERENCE: 023070-079810DS Factor Binding Protein From Legume Roots

CURRENT APPLICATION NUMBER: US/09/129,112

PRIOR FILING DATE: 1998-08-04

PRIOR FILING DATE: US 08/907,226

NUMBER OF SEQ ID NOS: 19

SOFTWARE: Patent Ver. 2.1

SEQ ID NO 3

LENGTH: 6265

TYPE: DNA

ORGANISM: Dolichos biflorus

FEATURE:

NAME/KEY: genomic sequence of NBP46 (DB46)

LOCATION: (633)..(944)

NAME/KEY: Intron

LOCATION: (945)..(1022)

NAME/KEY: exon

LOCATION: (1023)..(1151)

NAME/KEY: Intron

LOCATION: (1152)..(1559)

NAME/KEY: exon

LOCATION: (1560)..(1616)

NAME/KEY: Intron

LOCATION: (1617)..(1697)

NAME/KEY: exon

LOCATION: (1698)..(1790)

US-09-129-112-3

Query Match
Best Local Similarity 42.5%; Pred. No. 0.94;
Matches 206; Conservative 0; Mismatches 279; Indels 0; Gaps 0;

QY 726 AAAAAAAGACATTTATCATTAATACATCTTATAGAGCCCTAATTTCTTATCTGAG 785
DB 2790 AAACGTAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2731
QY 786 GCATGCTCTCTCCAGAAAGATACCTATATTTTGGTACTTATCGAATCCAGATGC 845
DB 2730 ACTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2671

QY 846 GATACCTCCCTGTCACAGATGACAAATGTGTTTCATTAAGTGGAGAGAAAACACATTT 905
DB 2670 AAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2611
QY 906 TAAAGTACTTTTGGGAGACTGATTTGATTAATTAATTAATTAATTAATTAATTAATTA 965
DB 2610 TTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2551

QY 966 AAAAAACCTCCACCTTACCTGTCATTTATATCCCTAGTCCAAAGTTAATAT 1025
DB 2550 CAAAGAACCTTAATTTCTTTTGAACATATCCATGTTTGAATCCATCTTAAATTTTC 2491

QY 1026 CTATTTCTGATATATGCTTTTATACCAAGAGCCCTATGACGAGTCCAGAACCA 1085
DB 2490 ATCATTTGCTTCTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 2431

QY 1086 CTATACGACACATACATACCAACATACCAATATACGTTTATATATCATAGTAA 1145
DB 2430 GCGACCCCTAAAGAACACCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2371

QY 1146 CTGACGACATGATTTATGAGCTTGTGATGTCGAATACGATTCATCATCATTTCTCA 1205
DB 2370 AATTAATATCTGCTGATTCATTCGACATTAATTAATTAATTAATTAATTAATTAATTA 2311

QY 1206 TCTTT 1210
DB 2310 CATTT 2306

RESULT 3
US-09-385-982-133
Sequence 133, Application US/09385982

Patent No. 6262334

GENERAL INFORMATION:

APPLICANT: ENDEGE, WILSON O., ET AL.

TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

FILE REFERENCE: CCDNA-260XX

CURRENT APPLICATION NUMBER: US/09/385,982

EARLIER FILING DATE: 1999-08-30

EARLIER FILING DATE: 09/328,111

EARLIER FILING DATE: 1999-06-08

EARLIER APPLICATION NUMBER: 60/117,393

EARLIER FILING DATE: 1999-01-27

EARLIER APPLICATION NUMBER: 60/098,639

EARLIER FILING DATE: 1998-08-31

NUMBER OF SEQ ID NOS: 544

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 133

LENGTH: 588

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)..(588)

OTHER INFORMATION: n - A,T,C or G

US-09-385-982-133

Query Match
Best Local Similarity 41.3%; Pred. No. 0.57;
Matches 95; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 35 TGTCTATGTACCAACAGTGAATTTATACAGTGTATCAATGTCTGGAATTTTGCACCTGC 94
DB 96 TCGGCNAGGTAACAGAGGTCAGATGTCNNGTGACAGAGACAGAAATCTGGCCNCA 155
QY 95 CATAGGAATGTTAAGTACTTGTGCTGGAATTTATCAACTTTGTAGTAACACTTGA 154
DB 156 CATTGANGTANNAANAAGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 215
QY 155 AGTTTACGATGAGGAGGGAATTTTGAAGCCCTTAAGCTTAACAAATTAATCAGATCT 214

APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereo
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-21
Query Match 2.8%; Score 37.2; DB 2; Length 246240;
Best Local Similarity 84.0%; Pred. No. 8.5;
Matches 42; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 912 AACTTTTGGAGACTGATTGAGTAATAATAAACTGCTCCCTTA 961
DB 154422 AATTCACAGAGACTGATTGAGTAATAATAAACTGCTCCCTTA 154471
RESULT 7
US-08-724-394A-22
Sequence 22, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kironmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereo
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-22
Query Match 2.8%; Score 37.2; DB 2; Length 246240;
Best Local Similarity 84.0%; Pred. No. 8.5;
Matches 42; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 912 AACTTTTGGAGACTGATTGAGTAATAATAAACTGCTCCCTTA 961
DB 154422 AATTCACAGAGACTGATTGAGTAATAATAAACTGCTCCCTTA 154471
RESULT 8
US-09-004-838-91/c
Sequence 91, Application US/09004838
Patent No. 6350933
GENERAL INFORMATION:
APPLICANT: Michelmore, Richard W.
APPLICANT: Shen, Kathy
APPLICANT: Meyers, Blake
TITLE OF INVENTION: Procedures and Materials for
TITLE OF INVENTION: Confering Pest Resistance in Plants
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,838
FILING DATE: 09-JAN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/781,734
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-0788100S
TELECOMMUNICATION INFORMATION:

RESULT 11
 US-09-234-827B-1/C
 ; Sequence 1, Application US/09234827B
 ; Patent No. 6448471
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: Puzio, Piotr S.
 ; APPLICANT: Grudler, Florian M.W.
 ; TITLE OF INVENTION: Nematode feeding structure specific gene and its

;; TITLE OF INVENTION: Application to produce nematode resistant plants
;; FILE REFERENCE: U-012084-2
;; CURRENT APPLICATION NUMBER: US/09/234,827B
;; CURRENT FILING DATE: 1999-01-21
;; PRIOR APPLICATION NUMBER: US 60/072,142
;; PRIOR FILING DATE: 1998-01-22
;; NUMBER OF SEQ ID NOS: 32
;; SOFTWARE: Patentln Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 8302
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (4038)..(4256)
;; NAME/KEY: CDS
;; LOCATION: (4807)..(5604)
;; NAME/KEY: CDS
;; LOCATION: (6777)..(6827)
;; NAME/KEY: CDS
;; LOCATION: (6919)..(7935)
US-09-234-827B-1

Query Match 2.7%; Score 36.2; DB 4; Length 8302;
Best Local Similarity 51.2%; Pred. No. 4.5;
Matches 11; Conservative 0; Mismatches 103; Indels 3; Gaps 1;

QY 683 CTCGATTTCTTTGTTTCTTTTAAAGACAGAGCTGTCTCAAAAAAGACATTGA 742
DB 805 CTTTATGTTGGATTTTATTAAGACAGACCGCAAAACATCAATACACTTA 746
QY 743 TCATTATACATCTATTTAGCCCTTATCTTATCGAAGGACGTTTCTTTT 802
DB 745 ATTTTATATCATTTTATCAATTTAATTAATTTGCTGTA--TTATTTTCAATTTAA 689
QY 803 AAACAGTTAGTACTGATGTCACAGACAAATTTTCTGATCAGATAGTCCCTGTCAAC 862
DB 688 AATAATTAATATTTAAATGTCCTCAATAAATAATGATCGGTTTATGTTTACC 629
QY 863 AGTACCAATGTGTTTCTTAAGGCGGAGAAACA 899
DB 628 TGTAAAGACTCTGTTTATCAAAAAAGAAAAA 592

RESULT 12
US-09-424-283-6
Sequence 6, Application US/09424283
Patent No. 6437219
GENERAL INFORMATION:
APPLICANT: Grimes, et al.
TITLE OF INVENTION: Sucrose binding proteins
FILE REFERENCE: 4630-50206
CURRENT APPLICATION NUMBER: US/09/424,283
CURRENT FILING DATE: 1999-11-19
PRIOR APPLICATION NUMBER: PCT/US98/10465
PRIOR FILING DATE: 1998-05-21
PRIOR APPLICATION NUMBER: US 60/047,568
PRIOR FILING DATE: 1997-05-22
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 6
LENGTH: 3718
TYPE: DNA
ORGANISM: Glycine max
US-09-424-283-6

Query Match 2.7%; Score 36; DB 4; Length 3718;
Best Local Similarity 47.0%; Pred. No. 3.8;
Matches 11; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 983 TTTACTGTCTATTATATCCCTAGTCCAAAGTAAATTAATTTCTGATATG 1042
DB 3257 TGTAGCACTATTGATTTTCTTCCACATTAATGAGTGAATCAGTTAGAGAA 3316

QY 1043 CTTTATACCAAGAGCCCTTATCAGCCAGTTCAGAACACCTATAGCAGAACATA 1102
DB 3317 ATATTAAAAAATAATTAATTAAGGAAGAGCATTAATTAAGATACGAACCTC 3376
QY 1103 CCAACCAATACCAACATTAATTAATCATCAGTACGACGACATGATTAAT 1162
DB 3377 ACAATGATAGACCAATTAAGACATTTATTTCTTCAAAATTAAGAAAGCTTTT 3436
QY 1163 TGAGGCTTGATTTGGCAATATACGACTCTACATTCATTTCAATCTTCAATACAT 1218
DB 3437 TTAACATATATACATTAATCATATATTAATTTTATTAATTTTAAACTT 3492

RESULT 13
US-09-009-913-1
Sequence 1, Application US/09009913
Patent No. 6087485
GENERAL INFORMATION:
APPLICANT: Arys Pharmaceuticals, Inc.
TITLE OF INVENTION: Asthma Related Genes
NUMBER OF SEQUENCES: 339
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Ave, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/009,913
FILING DATE: 21-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sherwood, Pamela J
REGISTRATION NUMBER: 36,677
REFERENCE/DOCKET NUMBER: SEQ-4P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3231
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 72928 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-09-009-913-1

Query Match 2.7%; Score 35.8; DB 3; Length 72928;
Best Local Similarity 54.1%; Pred. No. 13;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 708 AGACAGAGCTCTGTCTCAAAAAAGACATTATATCAATCTTATTTAGAGCC 767
DB 68237 AGAGTGAAGACTCTGTCAAAAAATTAATAAATAAATACTGTATTAATCTTTTG 68296
QY 768 CTAATTTCTTCTGAGAGGACCTGTTTCTTAAACGTTAGTACGATGCAACA 827
DB 68297 CACCAACATATATGATATATCAACATTTATTTTAAAGATTTTGACATTTGCTTTTAA 68356
QY 828 GACAAATATTTCTGA 842
DB 68357 TATAAATTTTAA 68371

RESULT 14
US-09-134-001C-920/C
Sequence 920, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 920
LENGTH: 3780
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-920

Query Match
Best Local Similarity 50.6%; Pred. No. 4.9; Length 3780;
Matches 86; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 273 TTTGAGAGTACTTTCTTTTGACGAAATTTTCATTCCTGCGCATTCCTATATTCCTCC 332
DB 3049 TTTGACAGACAGCTTTTGAATAATCTTTAGTCCATACCTGTTGATTTCTCAAT 2990
QY 333 CTTTATAGGAGCATTCGATTTCTCTTTGTGGGAATGCCCATTAAGCATTTTCAG 392
DB 2989 CTTTATGACTTTTGTGATTAATTTGCGCTTTGACAGAGATATATTCAGAAATGACAT 2930
QY 393 ATCTTTGATGTGACATTAATGCCATTAATGTAATGCCGTTATTTGTGTGA 442
DB 2929 CCGCTACTTTCTCGCTGATCTATCTCTGCGCTATCATTTTAATGAA 2880

RESULT 15
US-08-487-826B-13/C
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chituls, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08-487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned

REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match
Best Local Similarity 45.0%; Pred. No. 8.8; Length 19124;
Matches 134; Conservative 0; Mismatches 164; Indels 0; Gaps 0;
QY 903 TTTTAAGTAACTTTTGGGAGACTGATTTGATATATAAAGCTGCTCCCTTAA 962
DB 6981 TTTTAAAAAAT 6922
QY 963 GAAAAAAGCCCTTCCACCTTTACTGTGCTATTAATATCCCTTAGTTCCAAAGTTAAT 1022
DB 6921 AAAAAAAT 6862
QY 1023 TATCTATTTCTGATATTTGCTTTATATACCAAGGCTTATACAGGCTTACACAGCA 1082
DB 6861 TATATATATATATTTTATATATATATATATATATATATATATATATATATAT 6802
QY 1083 CCAGTATACGACCAACATATACCAACATATACCAATATATATATATATATATAT 1142
DB 6801 AATACGAAATATCAACAATATATATATATATATATATATATATATATATATAT 6742
QY 1143 TACAGCAGAGCAT 1200
DB 6741 TAACATTAAT 6684

Search completed: June 27, 2003, 16:19:53
Job time: 97 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 14:32:30 ; Search time 223 Seconds
(without alignments)
8860.093 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 1331
Sequence: 1 ctatgtttttagaacaag.....gcctttcaatgtactctg 1331

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104.6	7.9	438	US-09-924-340-5	Sequence 5, Appl1
2	104.6	7.9	438	US-09-992-600A-5	Sequence 5, Appl1
3	104.6	7.9	438	US-10-000-489-5	Sequence 5, Appl1
4	104.6	7.9	438	US-10-000-986-5	Sequence 5, Appl1
5	86.6	5.5	387	US-09-918-995-18537	Sequence 18537, A
6	74.8	5.5	433	US-09-983-965-5821	Sequence 5821, Ap
7	46.8	3.5	3104	US-10-037-598-32	Sequence 32, Appl
8	46.8	3.5	513509	US-09-754-853A-4	Sequence 4, Appl1
9	43.2	3.2	65359	US-09-804-472-3	Sequence 3, Appl1
10	43.2	3.2	6055	US-10-091-504-1746	Sequence 1746, Ap
11	42.2	3.2	6055	US-09-764-869-1746	Sequence 1746, Ap
12	42.2	3.2	7809	US-10-091-504-1744	Sequence 1744, Ap
13	42.2	3.2	7809	US-10-091-504-1744	Sequence 1744, Ap
14	42.2	3.2	7809	US-09-764-869-1744	Sequence 1744, Ap
15	42.2	3.2	7809	US-09-764-869-1744	Sequence 1744, Ap
16	41.8	3.1	7934	US-10-239-676-154	Sequence 154, App
17	40.6	3.1	556	US-10-092-154-1685	Sequence 1685, Ap
18	40.6	3.1	556	US-09-764-847-1685	Sequence 1685, Ap
19	40.6	3.1	175561	US-10-017-721-3	Sequence 3, Appl1

C 20	40.4	3.0	17849	US-10-092-154-1315	Sequence 1315, Ap
C 21	40.4	3.0	17849	US-09-764-891-10139	Sequence 10139, A
C 22	40.4	3.0	17849	US-09-764-847-1315	Sequence 1315, Ap
C 23	40.4	3.0	17862	US-10-092-154-1313	Sequence 1313, Ap
C 24	40.4	3.0	17862	US-09-764-891-10138	Sequence 10138, A
C 25	40.4	3.0	17862	US-09-764-847-1313	Sequence 1313, Ap
C 26	40.4	3.0	11047	US-10-239-676-188	Sequence 188, App
C 27	39.6	3.0	151	US-09-864-761-22646	Sequence 22646, A
C 28	39.6	3.0	492	US-09-864-761-5885	Sequence 5885, Ap
C 29	39.6	3.0	2000	US-09-938-842A-4305	Sequence 4305, Ap
C 30	39.6	3.0	14147	US-10-172-086-52	Sequence 4205, Ap
C 31	39.6	3.0	44848	US-09-968-113-42	Sequence 52, Appl
C 32	39.6	3.0	44848	US-09-776-874A-42	Sequence 42, Appl
C 33	39.4	3.0	1691139	US-10-067-514-1	Sequence 1, Appl1
C 34	39.2	2.9	5867	US-10-239-676-37	Sequence 37, Appl
C 35	39.2	2.9	14147	US-10-172-086-52	Sequence 52, Appl
C 36	39	2.9	378361	US-09-901-136-3	Sequence 3, Appl
C 37	38.6	2.9	5236	US-09-925-297-363	Sequence 363, App
C 38	38.6	2.9	6265	US-09-129-112-3	Sequence 3, Appl
C 39	38.6	2.9	10758	US-10-044-090-61	Sequence 61, Appl
C 40	38.4	2.9	472	US-09-918-995-375	Sequence 375, Ap
C 41	38.2	2.9	2000	US-09-938-842A-3631	Sequence 3631, Ap
C 42	38.2	2.9	2286	US-10-228-796-3	Sequence 3, Appl1
C 43	38.2	2.9	2286	US-09-191-687B-3	Sequence 97, Appl
C 44	38.2	2.9	7906	US-10-239-676-97	Sequence 133, App
C 45	38	2.9	588	US-09-871-161-133	

ALIGNMENTS

RESULT 1
US-09-924-340-5
Sequence 5, Application US/09924340
Publication No. US20030027248A1
GENERAL INFORMATION:
APPLICANT: Bejani, Stephanie
TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
FILE REFERENCE: 91. US2. REG
CURRENT APPLICATION NUMBER: US/09/924,340
CURRENT FILING DATE: 2001-08-06
PRIORITY APPLICATION NUMBER: US 60/305,456
PRIORITY FILING DATE: 2001-07-13
PRIORITY APPLICATION NUMBER: US 60/302,277
PRIORITY FILING DATE: 2001-06-29
PRIORITY APPLICATION NUMBER: US 60/298,698
PRIORITY FILING DATE: 2001-06-15
PRIORITY APPLICATION NUMBER: US 60/293,574
PRIORITY FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: Jpatent
SEQ ID NO 5
LENGTH: 438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..83
NAME/KEY: CDS
LOCATION: 84..317
NAME/KEY: 3'UTR
LOCATION: 318..438
NAME/KEY: polyA_signal
LOCATION: 397..402
NAME/KEY: polyA_site
LOCATION: 423..438
US-09-924-340-5

Query Match 7.9% Score 104.6; DB 9; Length 438;
Best Local Similarity 88.3%; Pred. No. 7.6e-17;
Matches 113; Conservative 0; Mismatches 15; Indels 0;

PRIOR APPLICATION NUMBER: US 09/924,340
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: PCT/IB01/01715
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/305,456
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/302,277
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/298,698
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 60/293,574
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 112
SOFTWARE: Jpatent
SEQ ID NO 5
LENGTH: 438
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..83
NAME/KEY: CDS
LOCATION: 84..317
NAME/KEY: 3'UTR
LOCATION: 318..438
NAME/KEY: polyA_signal
LOCATION: 397..402
NAME/KEY: polyA_site
LOCATION: 423..438
US-10-000-986-5

Query Match
Best Local Similarity 88.3%; Pred. No. 7 6e-17;
Matches 113; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1204 CACCTTCATACATATACACTACTACCACTTTTGTAGATCATCTAGACCAATGCG 1263
DB 192 CACATTCATACATATACACTACTACCACTTTTGTAGATCATCTAGACCAATGCG 251
QY 1264 AATGTAAACCCATATATATAGATGATCTTGTGTTCCAGTACTTGCCTTTTCAAT 1323
DB 252 AATGAAAACACATATATATAGTACTTGTGTTCCAGTACTTGCCTTTTCAAT 311
QY 1324 GTCACTTG 1331
DB 312 GTCACTTG 319

RESULT 5
US-09-918-995-18537/c
Sequence 18537, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: fastseq for Windows Version 3.0
SEQ ID NO 18537
LENGTH: 387
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(387)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-18537

Query Match
Best Local Similarity 6.5%; Score 86.6; DB 10; Length 387;
Matches 122; Conservative 0; Mismatches 34; Indels 5; Gaps 1;

QY 812 AGTACTGATGTCACAGACAAATATTTCTGATCAGATAGTCCCTGTCACAGTACGAA 871
DB 345 AGAAGACATGAAATTTTCCAAATATTTCCGATCAGAGATC-----ACAAGACAGCA 291
QY 872 TGTGTTTCATTAAGTCGGAAGAAACAGCATTTTAAGTACTTTTGGGAGACTGATT 931
DB 290 TGTGTTTCATTAAGTCGGAAGAAACAGCATTTTAAGTACTTTTGGGAGACTGATT 231
QY 932 TGAATATATTAATAACTCTGCTCCTTAAGAAAAA 972
DB 230 TGAATATATTAATAACTCTGCTCCTTAAGAAAAA 190

RESULT 6
US-09-983-965-5821
Sequence 5821, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Rao, Nengbing
APPLICANT: Byatt, John C.
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983,965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 5821
LENGTH: 433
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
OTHER INFORMATION: Clone ID: 60-LIB34-039-Q1-E1-G8
US-09-983-965-5821

Query Match
Best Local Similarity 5.6%; Score 74.8; DB 10; Length 433;
Matches 142; Conservative 0; Mismatches 62; Indels 10; Gaps 2;

QY 482 TTGATTTTGTGTCACCTTACCAATAGCCTTTTGAATGACGAAAGTGAACAGAGAA 541
DB 4 TGGGTTTCTGCTTACTCTCTCAAGAGCCTTTGAATGACGAG-----ATGTCACAGA 57
QY 542 GAGCATGTCGCAAGAAAGATAGCTAATATTTTGTACTTATCTGAATCCAGAA 601
DB 58 GAAAGAACTGTCAGAAAGAGTATGATTTTGTGGAAGCTTGAATCCAGAAATCCAGAA 117
QY 602 TGTGCTTCCCTGAGGTGTTTCTTATGATGATCTGATGAATCCCTCGGAG 661
DB 118 TGTGCTTCTGAGGTGTTTCTTATGATGATCTGATGAATCCCTCGGAG 661
QY 662 CACAGACAGTTAGTAGAATCTCCATTTCTTTG 695
DB 174 GATAGCCAGCAGAGTAGAATCTGATTTCTTTG 207

RESULT 7
US-10-037-598-32
Sequence 32, Application US/10037598
Patent No. US20020157143A1
GENERAL INFORMATION:
APPLICANT: Monsanto Co
APPLICANT: Conclido, Vergel
APPLICANT: Delaney, Xavier

792 TTTT TTTT TTTT AA CAGT TAAGT ACTG ATGT CACA CAGACA AATAT TTTCTG ATCAGATAGT 851

FILE REFERENCE: PC007C1

Page 4

```
TTCAT 3019  
      |  
AAAGT 911  
TTATA 3019  
AAAAA 971  
IIIII  
AAAAA 3020  
  
025  
02075
```

CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1746
LENGTH: 6055
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1746

Query Match
Best Local Similarity 45.8%; Score 42.2; DB 9; Length 6055;
Pred. No. 3.2;
Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 937 AATATAAATCTGCTGCTCCCTTAAGAAAAAACCCCTTCACCTTACTGTGTCATT 996
DB 835 AAAGAAAGATGATCTTTCTTCTAGTTAAATGAATGTAACCCGTAAGAAATGTTTAA 894
QY 997 TATATCCCTTATGTTCCAAAGTTAATATCTTATTCGATATTCCTTTTATACCAAG 1056
DB 895 AAGACTGCCAGACTCTTTAAGGCTTTAAGCTTATGTTATTTATCTGTTTATTTACT 954
QY 1057 AGCCTTATCAGCAGCTTCAGAACACACACTATACGCAACATACCAACCAATACC 1116
DB 955 ATAGAAATAGATTAATTAAGAAAGATGAATTAATTAATTTTGTACATTTT 1014
QY 1117 AACAAATATAGCTTTAATATCATCATGACTGACGACATGATTTAGGCTTGATTTG 1176
DB 1015 AAATTAATATGATTTCTTATATGATTAATTAATTAATTAATTTTATTTTAA 1074
QY 1177 CAAATAGCACTTACATTCATATTCCTTTTACATCATCATACACACTACACCT 1236
DB 1075 GTAAACAACTTTAAGTACAGCCTTATATTTAAATTTCTGTAGGCTAGAAATCAGATT 1134
QY 1237 TTGTGNAGATCA 1248
DB 1135 AGTGCCATAGCA 1146

RESULT 11

US-09-764-869-1746
Sequence 1746, Application US/09764869
Patent No. US20020061521A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007
CURRENT APPLICATION NUMBER: US/09/764,869
CURRENT FILING DATE: 2001-01-17
Prior Application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1746
LENGTH: 6055
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-869-1746

Query Match
Best Local Similarity 45.8%; Score 42.2; DB 10; Length 6055;
Pred. No. 3.2;
Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 937 AATATAAATCTGCTGCTCCCTTAAGAAAAAACCCCTTCACCTTACTGTGTCATT 996
DB 835 AAAGAAAGATGATCTTTCTTCTAGTTAAATGAATGTAACCCGTAAGAAATGTTTAA 894
QY 997 TATATCCCTTATGTTCCAAAGTTAATATCTTATTCGATATTCCTTTTATACCAAG 1056
DB 895 AAGACTGCCAGACTCTTTAAGGCTTTAAGCTTATGTTATTTATCTGTTTATTTACT 954
QY 1057 AGCCTTATCAGCAGCTTCAGAACACCACTATACGCAACATACCAACCAATACC 1116
DB 955 ATAGAAATAGATTAATTAAGAAAGATGAATTAATTAATTTTGTACATTTT 1014

QY 1117 AACATATAGCTTTAATATCATCATGACTGACGACATGATTTAGGCTTGATTTG 1176
DB 1015 AAATTAATATGATTTCTTATATGATTAATTAATTAATTAATTTTATTTTAA 1074
QY 1177 CAAATAGCACTTACATTCATATTCCTTTTACATCATCATACACTACACCT 1236
DB 1075 GTAAACAACTTTAAGTACAGCCTTATATTTAAATTTCTGTAGGCTAGAAATCAGATT 1134
QY 1237 TTGTGNAGATCA 1248
DB 1135 AGTGCCATAGCA 1146

RESULT 12

US-10-091-504-1744
Sequence 1744, Application US/10091504
Publication No. US20030059908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
CURRENT FILING DATE: 2002-03-07
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1744
LENGTH: 7809
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1744

Query Match
Best Local Similarity 45.8%; Score 42.2; DB 9; Length 7809;
Pred. No. 3.7;
Matches 143; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 937 AATATAAATCTGCTGCTCCCTTAAGAAAAAACCCCTTCACCTTACTGTGTCATT 996
DB 2277 AAAGAAAGATGATCTTTCTTCTAGTTAAATGAATGTAACCCGTAAGAAATGTTTAA 2336
QY 997 TATATCCCTTATGTTCCAAAGTTAATATCTTATTCGATATTCCTTTTATACCAAG 1056
DB 2337 AAGACTGCCAGACTCTTTAAGGCTTTAAGCTTATGTTATTTACTCTTTTATTTACT 2396
QY 1057 AGCCTTATCAGCAGCTTCAGAACACCACTATACGCAACATACCAACCAATACC 1116
DB 2397 ATAGAAATAGATTAATTAAGTGAAGAAAGATTAATTAATTTTGTACATTTT 2456
QY 1117 AACAAATATAGCTTTAATATCATCATGACTGACGACATGATTTAGGCTTGATTTG 1176
DB 2457 AAATTAATATGATTTCTTATATGATTAATTAATTAATTAATTTTATTTTAA 2516
QY 1177 CAAATAGCACTTCTATCATCATATTCCTTTTACATCATCATACACTACTACACT 1236
DB 2517 GTAAACAACTTTAAGTACAGCCTTATATTTAAATTTCTGTAGGCTAGAAATCAGATT 2576
QY 1237 TTGTGNAGATCA 1248
DB 2577 AGTGCCATAGCA 2588

RESULT 13

US-10-091-504-1745
Sequence 1745, Application US/10091504
Publication No. US20030059908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442

QY	937	AATATATAAAACCTGGCTCCCTTAGAGAAAAAAACCCTTCCACCTTACGTGTCATT	996
Db	2277	AAAGAAAGATCTTTTCTAGTTAATAGAAAGCTAACCGCTACTCGAAGAAAGTTTTTAA	2336
QY	997	TATATCCCTTAGTTCCAAAGTTATTTATCTTATTCGTGATATGCTTTATATACCAAG	1056
Db	2337	AAGACTGCAGACTCTTTAAAGCGTTTACCATTTGATTTTATCTGCTTTTATTTCTAACT	2396
QY	1057	AGCTTATATAGCCAGTTCGAGACACACCATATACGACACAAACCATCAACAACACATAC	1116
Db	2397	ATAGGAAATAGATTAATTACTGAGAAAGATGAATTAATAATTAGTTATTGTACAAATTTT	2456
QY	1117	AACAATATACGTTTATATATCATCATGATACGACGACGACATGATTTATGAGCGCTGATATGG	1176

QY	93	AAATAAATAACCTGTGCTCCCTTAAGAAAAAAACCCTTCACCTTACGTGTGCATT	996
Db	2276	AAAGAAAGATACCTTTTCTAGTAATGAAGTAAACCGCTACTCAGAAAGTTTTTAA	2333
QY	997	TATATCCCTTGTGTTCCAAAGTTAATATCTATTTCTGATATTCCTTTATACCAAG	1056
Db	2336	AAACATGCCAGACTCTTTAAGGGTTTAACATATGATTAATCTCTCTTTATTTTACCT	2392
QY	1057	AGCCTTATCAGCAGCTTCCAGAACACACATATACGACACACATACCAACCAATAC	1116
Db	2396	ATAGGAAATAGATATTTCTGAAAGAAAGATGAATTAATTAAGTTATTTGTACATTTT	2455
QY	1117	AACATATACGTTTTTAATATCATCAGTACTGCAGGACATGATTTATGAGCTTGATTGG	1176
Db	2456	AAATAAATATGTTTTCTTATATGTATAAATAAAAAACCGTATATGATTTTTTTTAA	2515
QY	1177	CAAAATACGCTCTCATCATATTTCTACATCTTTCATACCATATACACACTACTACACT	1236
Db	2516	GTAAAAACAACCTTTTACGTACAGCCTTTATTTTAAATCTGTAGGGCTAGAAATCAGAA	2575
QY	1237	TTTGTNAGATCA	1248
Db	2576	AGTGCATATGCA	2587

Search completed: June 27, 2003, 16:23:52
Job time : 229 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 12:58:01 ; Search time 1825 seconds
(without alignments)
11811.613 Million cell updates/sec

Title: US-09-830-244B-2

Perfect score: 1331

Sequence: 1 catgttttagatcaaga.....gcctttccaatgtcactgt 1331

Scoring table: IDENTITY_NUC

Gâpâp 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlmu:*
5: em_estlmu:*
6: em_estlmu:*
7: em_estlmu:*
8: em_estlmu:*
9: em_estlmu:*
10: gb_est1:*
11: gb_est2:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estlmu:*
16: em_estlmu:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	490	36.8	492	9	AI224097
C 2	480	36.1	518	9	AI694371
C 3	478.4	35.9	523	10	AM269813
C 4	405	30.4	888	12	BG121534
C 5	399	30.0	427	12	BF831772
C 6	398.4	29.9	417	12	BE767736

C 7	380.4	28.6	394	12	BG011356
C 8	369.4	27.8	403	12	BF831960
C 9	353	26.5	408	12	BG011354
C 10	352.8	26.5	392	12	BE29791
C 11	345.6	26.0	402	12	BG011351
C 12	344.8	25.9	395	12	BF96088
C 13	335.2	25.2	951	12	BF168510
C 14	333	25.0	396	12	BF832034
C 15	330.4	24.8	376	9	AA488304
C 16	322	24.2	386	14	BQ359015
C 17	309.2	23.2	352	14	BQ359019
C 18	293.2	22.0	324	12	BF986192
C 19	280.8	21.1	417	14	BQ359018
C 20	266.6	20.0	295	10	AM135014
C 21	266.6	19.7	279	12	BF195348
C 22	262.4	19.0	365	10	BE089961
C 23	252.6	18.8	517	12	BG188342
C 24	249.8	18.7	470	12	BG200817
C 25	248.8	18.5	404	9	AA376713
C 26	246.2	18.4	405	9	AA376719
C 27	245.2	18.2	519	10	AA951480
C 28	242.4	18.0	250	9	AA30950
C 29	239	17.9	410	14	T29650
C 30	238.4	17.7	375	9	AA376681
C 31	236	17.2	335	9	AA376714
C 32	228.6	16.7	257	12	BF832847
C 33	222	15.9	457	9	AA376682
C 34	211	15.3	306	10	BE163886
C 35	203.4	14.9	426	9	AA376680
C 36	197.8	14.6	302	9	AA376694
C 37	181.4	13.5	341	9	AA376695
C 38	179.2	13.2	329	14	BQ309679
C 39	175.6	12.7	321	12	BG205925
C 40	169.6	12.2	258	12	BF834202
C 41	163	12.1	314	10	AA409123
C 42	161.4	12.1	209	12	BF835534
C 43	161	11.7	173	12	BF088945
C 44	155.6	11.6	172	12	BF944749
C 45	155				

ALIGNMENTS

RESULT 1
LOCUS AI224097 492 bp MRNA linear EST 28-JAN-1999
DEFINITION g135a03.x1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:1858444
3' mRNA sequence.

ACCESSION AI224097
VERSION AI224097.1 GI:3806810
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-femail.nih.gov
This clone is available royalty-free through LNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 876 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 456.
Location/Qualifiers
1. 492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1858444"

FEATURES

source
1. 492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1858444"


```

RESULT 3
AM269813/c 523 bp mRNA linear EST 03-JAN-2000
LOCUS xV45f02.x1 Soares_NFL_T_GBC_SI Homo sapiens cDNA clone
DEFINITION IMAG:2816091.3', mRNA sequence.
ACCESSION AM269813
VERSION AM269813.1 GI:6656843
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 523)
AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40up from gibco
High quality sequence stop: 462.
FEATURES
source
1..523
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2816091"
/clone.lib="Soares_NFL_T_GBC_SI"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site:1: Not 1; Site:2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHF, and B-cell
NCI CGAP GCB1) were mixed, and 5S circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
1.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731395. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 177 a 109 c 85 g 132 t
ORIGIN
Query Match 35.9%; Score 478.4; DB 10; Length 523;
Best Local Similarity 99.8%; Pred. No. 4.3e-102;
Matches 479; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTATGTTTGAATCAAGATGACCGGTAAAGCTGCTCATGTACCAACGTAATTT 60
DB 481 CTATGTTTGAATCAAGATGACCGGTAAAGCTGCTCATGTACCAACGTAATTT 422
QY 61 ACAGTGTTCACAAATGCTGGAATTTGCACTGCGCATAGGGAATGTTAAAGCTTGGC 120
DB 421 ACAGTGTTCACAAATGCTGGAATTTGCACTGCGCATAGGGAATGTTAAAGCTTGGC 362
QY 121 TGAATTTATCAGACTGTGAGTAACAAGTTGAAGTTAGCAGATAGGGGGAATTG 180
DB 361 TGAATTTATCAGACTGTGAGTAACAAGTTGAAGTTAGCAGATAGGGGGAATTG 302
QY 181 AGCCCCCTAAGGCTAAACAAATATCATGATCTGAGATAGTGGCTAATGTGCTCCCA 240
DB 301 AGCCCCCTAAGGCTAAACAAATATCATGATCTGAGATAGTGGCTAATGTGCTCCCA 242
QY 241 GGCCTAATTTGGGAACAGTTTTCGATGCTTGTAGAGTACTTTCTTTTACAGAAA 300
DB 241 GGCCTAATTTGGGAACAGTTTTCGATGCTTGTAGAGTACTTTCTTTTACAGAAA 182
QY 301 TTTTCATTTCTGCTTGCATGCTATATTTCTCCCTTTATAGAGCCTATTGATTTCTTC 360
DB 181 TTTTCATTTCTGCTTGCATGCTATATTTCTCCCTTTATAGAGCCTATTGATTTCTTC 122

```

```

QY 361 TTTTGGGAATGTCATTCATTTTCAGATCTTTGATGTCAGTAAATGCAATAT 420
DB 121 TTTTGGGAATGTCATTCATTTTCAGATCTTTGATGTCAGTAAATGCAATAT 62
QY 421 TGGTAATGCCGTTATTTGTAATACAGCATAGTTAAATTAAGTGTACAGTAATCTACA 480
DB 61 TGGTAATGCCGTTATTTGTAATACAGCATAGTTAAATTAAGTGTACAGTAATCTACA 2
RESULT 4
BG121534 888 bp mRNA linear EST 30-JAN-2001
LOCUS 602352872P1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4451201.5',
DEFINITION mRNA sequence.
ACCESSION BG121534
VERSION BG121534.1 GI:12615043
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 888)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM10237 row: p column: 18
High quality sequence start: 2
High quality sequence stop: 680.
FEATURES
source
1..888
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4451201"
/clone.lib="NIH_MGC_90"
/lisuse-type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-Sport6; Site:1: NotI;
Site:2: SalI; cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC library."
BASE COUNT 240 a 161 c 198 g 289 t
ORIGIN
Query Match 30.4%; Score 405; DB 12; Length 888;
Best Local Similarity 89.8%; Pred. No. 7.9e-85;
Matches 468; Conservative 0; Mismatches 50; Indels 3; Gaps 3;
QY 1 CTATGTTTGAATCAAGATGACCGGTAAAGCTGCTCATGTACCAACGTAATTT 60
DB 284 CTATGTTTGAATCAAGATGACCGGTAAAGCTGCTCATGTACCAACGTAATTT 343
QY 61 ACAGTGTTCACAAATGCTGGAATTTGCACTGCGCATAGGGAATGTTAAAGCTTGGC 120
DB 344 ACAGTGTTCACAAATGCTGGAATTTGCACTGCGCATAGGGAATGTTAAAGCTTGGC 403
QY 121 TGAATTTATCAGACTGTGAGTAACAAGTTGAAGTTAGCAGATAGGGGGAATTG 180
DB 404 TGAATTTATCAGACTGTGAGTAACAAGTTGAAGTTAGCAGATAGGGGGAATTG 463
QY 181 AGCCCCCTAAGGCTAAACAA-ATATCATGATCTGAGATAGTGGCTAATGTGCTCC 239
DB 464 AGCCCCCTAAGGCTAAACAAATATCATGATCTGAGATAGTGGCTAATGTGCTCC 523

```

QY 240 AGCCCTAATTTGGAGACAGTCTTTCTGATGCTTGAGAGTACTTCTTTGACAGAA 299
 Db 524 AGG-CTAATTTGGAGACAGTCTTTCTGATGCTTGAGAGTACTTCTTTGACAGAA 582
 QY 300 ATTTTCAATTCCTGCTTCCCATTCGTATATTCCTTTATAGAGGACCATTTGCTTTC 359
 Db 583 ATTTTCAATTCCTGCTTCCCATTCGTATATTCCTTTATAGAGGACCATTTGCTTTC 641
 QY 360 CTTTGTGGGAATGTCCTATAGCATTTTTCAGATCTTTTGTGACACTAATGCCATTA 419
 Db 642 CTTTGTGGGAATGTCCTATAGCATTTTTCAGATCTTTTGTGACACTAATGCCATTA 701
 QY 420 TTGCTAATGCTGCTTATGCTGATACAGCATAGTAAATACCTTTCAGTAAATCCTAC 479
 Db 702 ATGGTAAATGCTGCTTATGCTGATACAGCATAGTAAATACCTTTCAGTAAATCCTAC 761
 QY 480 ACTTGATTTGCTGACCTCTACCAATAGCCTTTTGAATGA 520
 Db 762 ACCTGGGCTTTTGGGCGCTTACAAAGCTTTGATGACGGA 802

RESULT 5
 LOCUS BF831772/c 427 bp mRNA linear EST 13-JAN-2001
 DEFINITION PM3-HT0909-151000-009-e12 HT0909 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF831772
 VERSION BF831772.1 GI:12179873
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 427)
 AUTHORS Dias Neto, E., Garcia Correa, R., Veijovskij-Almeida, S., Bioness, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM3&t2=PM3-HT0909-
 151000-009-e12&t3=2000-10-15&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 18
 Location/Qualifiers
 1..427
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="HT0909"
 /dev_stage="Adult"
 /note="Organ: head, neck; Vector: puc18; Site: 1; Smat:
 Site: 2; Smat: A mini-library was made by cloning products
 derived from ORESSES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 146 a 95 c 72 g 113 t 1 others

Query Match 30.0%; Score 399; DB 12; Length 427;
 Best Local Similarity 99.8%; Pred. No. 2, 1e-83;
 Matches 399; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 42 TGTACCAACGTAATTTTACAGTGTGTTTACAAATGCTGGAATTTTGCACCTCCATAGGC 101
 Db 427 TGTACCAACGTAATTTTACAGTGTGTTTACAAATGCTGGAATTTTGCACCTCCATAGGC 368
 QY 102 AATGTTAAGGTAATCTGGCGTGAATTTATCAAGTGTGTAAGTAAGTAAGTGAAGTTAG 161
 Db 367 AATGTTAAGGTAATCTGGCGTGAATTTATCAAGTGTGTAAGTAAGTAAGTGAAGTTAG 308
 QY 162 CAGATAGGAGGGAATATTTAGAGCCCTTAAGGCTTAACAAATTAATACAGTATGATGATAG 221
 Db 307 CAGATAGGAGGGAATATTTAGAGCCCTTAAGGCTTAACAAATTAATACAGTATGATGATAG 248
 QY 222 TGGCTAATGTGCTCCCGACGCTTAATTTGGAGACAGTTTTCGTGATGCTTTGAGAGAG 281
 Db 247 TGGCTAATGTGCTCCCGACGCTTAATTTGGAGACAGTTTTCGTGATGCTTTGAGAGAG 188
 QY 282 TACTTCTTTGACAGAAATTTTCAATTCGCTTGCATTCATATTCCTTATAGC 341
 Db 187 TACTTCTTTGACAGAAATTTTCAATTCGCTTGCATTCATATTCCTTATAGC 128
 QY 342 AGCCATGATTTCTTCTTCTTTTGGGAAATGCCATTCATTTGACATCTTTTGA 401
 Db 127 AGCCATGATTTCTTCTTCTTCTTTTGGGAAATGCCATTCATTTGACATCTTTTGA 68
 QY 402 TGTGACATTAATCCATTTATGCTATGCTTATGCTGTA 441
 Db 67 TGTGACATTAATCCATTTATGCTATGCTTATGCTGTA 28

RESULT 6
 LOCUS BE767736/c 417 bp mRNA linear EST 20-SEP-2000
 DEFINITION RC3-GN0042-100800-011-e04 GN0042 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE767736
 VERSION BE767736.1 GI:10221394
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 417)
 AUTHORS Dias Neto, E., Garcia Correa, R., Veijovskij-Almeida, S., Bioness, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC3-GN0042-100
 800-011-e04&t3=2000-08-10&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 37
 Location/Qualifiers
 1..417


```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0042"
/dev_stage="Adult"
/note="Organ: Placenta-normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the puc18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT      143 a      92 c      71 g      111 t
ORIGIN

```

```

Query Match      29.9%; Score 398.4; DB 12; Length 417;
Best Local Similarity 99.8%; Pred. No. 2.9e-83;
Matches 399; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 42 TGTACCAACGTAAGATTACAGTCTTACAAATGTGCAATTTGGCAGTCCATAGG 101
DB 417 TGTACCAACGTAAGATTACAGTCTTACAAATGTGCAATTTGGCAGTCCATAGG 358
QY 102 AATGTTAGGTTACTGGCTGGAATTTTCAGACTTGTGAGTAAACAAGTTGAAGTTAG 161
DB 357 AATGTTAGGTTACTGGCTGGAATTTTCAGACTTGTGAGTAAACAAGTTGAAGTTAG 298
QY 162 CAGATGAGGGGGAATATGAGGCCCTTAAGGCTAAACAATATATCATCTGAGATAG 221
DB 297 CAGATGAGGGGGAATATGAGGCCCTTAAGGCTAAACAATATATCATCTGAGATAG 238
QY 222 TGGCTAATGTGGTCCCGCAGGCTTAATTTGGGAACAGTTTTCCTGATGCTTTGAGAAG 281
DB 237 TGGCTAATGTGGTCCCGCAGGCTTAATTTGGGAACAGTTTTCCTGATGCTTTGAGAAG 178
QY 282 TACTTCTTTGACAGAAATTTTCATCTGCTTCCATGCTATATCTCCCTTATAGG 341
DB 177 TACTTCTTTGACAGAAATTTTCATCTGCTTCCATGCTATATCTCCCTTATAGG 118
QY 342 AGCCATTGAGATTTCTTCTTTGCTTGGGAATGTCCTTATGAGATTTTTCAGATCTTTGA 401
DB 117 AGCCATTGAGATTTCTTCTTTGCTTGGGAATGTCCTTATGAGATTTTTCAGATCTTTGA 58
QY 402 TGTGCACTAATGCCATTATTTGTAATGCCCTTATTTGCTGA 441
DB 57 TGTGCACTAATGCCATTATTTGTAATGCCCTTATTTGCTGA 18

```

```

RESULT 7
LOCUS BG011356 394 bp mRNA linear EST 24-JAN-2001
DEFINITION R01-GN0268-091200-011-g04 GN0268 Homo sapiens cDNA, mRNA sequence.
ACCESSION BG011356
VERSION BG011356.1 GI:12459473
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE 1 (bases 1 to 394)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

```

```

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

```

```

Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC1&t2=RC1-GN0268-
091200-011-g04&t3=2000-12-09&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
Location/Qualifiers
1. 394
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0268"
/dev_stage="Adult"
/note="Organ: Placenta-normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the puc18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT      137 a      86 c      67 g      104 t
ORIGIN

```

```

Query Match      28.6%; Score 380.4; DB 12; Length 394;
Best Local Similarity 99.2%; Pred. No. 4.9e-79;
Matches 393; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

```

```

QY 42 TGTACCAACGTAAGATTACAGTCTTACAAATGTGCAATTTGGCAGTCCATAGG 101
DB 394 TGTACCAACGTAAGATTACAGTCTTACAAATGTGCAATTTGGCAGTCCATAGG 335
QY 102 AATGTTAGGTTACTGGCTGGAATTTTCAGACTTGTGAGTAAACAAGTTGAAGTTAG 161
DB 334 AATGTTAGGTTACTGGCTGGAATTTTCAGACTTGTGAGTAAACAAGTTGAAGTTAG 275
QY 162 CAGATGAGGGGGAATATGAGGCCCTTAAGGCTAAACAATATATCATCTGAGATAG 221
DB 274 CAGATGAGGGGGAATATGAGGCCCTTAAGGCTAAACAATATATCATCTGAGATAG 215
QY 222 TGGCTAATGTGGTCCCGCAGGCTTAATTTGGGAACAGTTTTCCTGATGCTTTGAGAAG 281
DB 214 TGGCTAATGTGGTCCCGCAGGCTTAATTTGGGAACAGTTTTCCTGATGCTTTGAGAAG 155
QY 282 TACTTCTTTGACAGAAATTTTCATCTGCTTCCATGCTATATCTCCCTTATAGG 341
DB 154 TACTTCTTTGACAGAAATTTTCATCTGCTTCCATGCTATATCTCCCTTATAGG 95
QY 342 AGCCATTGAGATTTCTTCTTTGCTTGGGAATGTCCTTATGAGATTTTTCAGATCTTTGA 401
DB 94 AGCCATTGAGATTTCTTCTTTGCTTGGGAATGTCCTTATGAGATTTTTCAGATCTTTGA 37
QY 402 TGTGCACTAATGCCATTATTTGTAATGCCCTTATTTGCTGA 437
DB 36 TGTGCACTAATGCCATTATTTGTAATGCCCTTATTTGCTGA 1

```

```

RESULT 8
LOCUS BF831960 403 bp mRNA linear EST 13-JAN-2001
DEFINITION PM3-HT0909-181000-010-g05 HT0909 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF831960
VERSION BF831960.1 GI:12180211
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 403)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

```


Qy	126	TTTATCAGCACTTGTGATGAACAAAGTTGAAGTTTAGCAGATGAGGGGGAATTATGAGGCC	181
Dd	80	TTTTACCAGACTTGTTGAGTAACAAGATTGAAGTTTACAGATGAGGGGGAATTATGAGGCC	139
Qy	186	CCTAAGGCGTAAACAAATAATCATCTATCTAGATAGTGGCTAATGTGGCTCCCGAGGCCT	241
Dd	140	CCTAAGGCTAAACAAATAATCATCTATCTAGATAGTGGCTAATGTGGCTCCCGAGGCCT	199
Qy	246	AATTTGGGACAGTTTTTCCTGATTTGGCTTGGAGAATGACTTTCTTTTGACAGAAATTTTC	307
Qy	200	AATTTGGGACAGTTTTTCCTGATTTGGCTTGGAGAATGACTTTCTTTTGACAGAAATTTTC	255
Dd	260	ATTTCGCTTGCATTTGCTATATTTCTCCCTTTATGAGGACGATGGATTTCTTTCCTTTAG	319
Qy	366	TGGGAATGTGCCATTTAGCATTTTTCAGATCTTTTGTATGGCACTAATGCGCATTTATTTGTA	425
Dd	320	TGGGAATGTGCCATTTAGCATTTTTCAGATCTTTTGTATGGCACTAATGCGCATTTATTTGTA	379
Qy	426	ATGCGGTTATTTGG 438	
Dd	380	ATGCGGTTATTTGG 392	
RESULT 11	BG011351/c	402 bp	MRA linear EST 24-JAN-2000
LOCUS	BG011351		
DEFINITION	RC1-GN0268-091200-011-b01 GN0268 Homo sapiens cDNA, mRNA sequence.		
ACCESSION	BG011351		
VERSION	BG011351.1	GI:12459463	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 402)		
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Negal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J		
TITLE	Spotgun sequencing of the human transcriptome with ORF expressed sequence tags		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)		
MEDLINE	20202663		
COMMENT	Contact: Simpson A.J.G. .Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel.: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC1&t2=RC1-GN0268- 091200-011-b01&t3=2000-12-09&t4=1) Seq primer: puc 18 forward High quality sequence start: 28 High quality sequence stop: 402. Location/Qualifiers 1..402 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_id="GN0268" /dev_stage="Adult" /note="Organ: placenta.normal; Vector: puc18; Site: 1; SmaI ; Site: 2; SmaI; A mini-library was made by cloning products derived from ORFES PCR (U.S. Letters Patent application No. 196,716; Inventor: Dr. David J. States")"		
FEATURES	source		

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: L1AM10228 Row: n Column: 16
High quality sequence stop: 675.

FEATURES
Source
Location/Qualifiers

1. 951
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="IMAGE:4447695"
/clone_id="NIH_MGC_89"
/tissue_type="hypernephroma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pCMV-Sport6; Site_1: Not; Site_2: SalI; cloned unidirectionally; oligo-dt primed. Average insert size 1.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."
BASE COUNT 281 a 221 c 196 g 253 t
ORIGIN

Query Match 25.2%; Score 335.2; DB 12; Length 951;
Best Local Similarity 95.8%; Pred. No. 2.1e-68;
Matches 366; Conservative 0; Mismatches 13; Indels 3; Gaps 2;

709 GACAGAGACTGTCTCAAAAAAGACATTATCATATTAACATTTATAGAGCCG 768
76 GACAGAGACTGTCTCAAAAAAGACATTATCATATTAACATTTATAGAGCCG 135
769 TATTTCTATCTGAAGGACGCTTTTATTTTAAACAGTTAAGTACTGATGCAAC 828
136 TATTTCTATCTGAAGGACGCTTTTATTTTAAACAGTTAAGTACTGATGCAAC 193
829 ACAAAATATTTCTGATCAGTATGTCCTCTGTCACAGTGAAGTGGTTTCATTAAGTG 888
194 ACAAAATATTTCTGATCAGTATGTCCTCTGTCACAGTGAAGTGGTTTCATTAAGTG 253
889 GGAAGAAAAAGACATTTTAAAGTAACTTTTGGGAGACTGATTCAGTAATATAAACT 948
254 GGAAGAAAAAGACATTTTAAAGTAACTTTTGGGAGACTGATTCAGTAATATAAACT 312
949 CTGGTCTCCCTTAAGAAAAAACCCCTTCACTTACTGTCATTTATATCCCTTAA 1008
313 CTGGTCTCCCTTAAGAAAAAACCCCTTCACTTACTGTCATTTATATCCCTTAA 372
1009 GTTCCAAAGTTAATATCTTATTTCTGATGATGCTTTTATACCAAGAGCTTATCAGC 1068
373 GTTCCAAAGTTAATATCTTATTTCTGATGATGCTTTTATACCAAGAGCTTATCAGC 432
1069 CAGTTCCAGAACACCACTATA 1090
433 CTTGTACTACAGTATCTTTA 454

RESULT 14
BF832034 396 bp mRNA linear EST 13-JAN-2001
DEFINITION PM3-HT0909-181000-011-h06 HT0909 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF832034
VERSION BF832034.1 GI:12180354
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 396)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Negal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsushima,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL
MEDLINE
COMMENT
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL:
(http://www.ludwig.org.br/scripts/gethtml2.pl?PL=PM3&ct=PM3-HT0909-181000-011-h06&f3=2000-10-18&f4=1)
Seq primer: puc 18 forward
High quality sequence stop: 87.

FEATURES
Source
Location/Qualifiers

1. 396
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="HT0909"
/dev_stage="Adult"
/note="Organ: head,neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 104 a 57 c 139 t
ORIGIN

Query Match 25.0%; Score 333; DB 12; Length 396;
Best Local Similarity 96.2%; Pred. No. 7e-66;
Matches 384; Conservative 0; Mismatches 10; Indels 5; Gaps 4;

42 TGTACCAAGCGTAATTTACAGTGTTCACAAATGTCGAATTTGGACGCGCATAGG 101
1 TGTACCAAGCGTAATTTACAGTGTTCACAAATGTCGAATTTGGACGCGCATAGG 60
102 AATGTTAAGCTTACTTGGCTGGAATTTATCAGACT--TGTAGTAACAAAGTTGAAGTTT 159
61 AATGTTAAGG-TACTTGGCTGGAATTTATGAAATTTGTAGTAACAAAGTTGAAGTTT 119
160 AACAGATGAGGGGGAATATTGAGGCCCCCTTAAGGCTAAACAAATATATCAGATCTGAGAT 219
120 AACAGATGAGGGGGAATATTGAGGCCCCCTTAAGGCTAAAC-AAATATATCAGATCTGAGAT 178
220 AGTGGCTATGTGGCTCCCGAGGCCCTAATTGGGAACAGTTTTCCTGATGCTTTGAGA 279
179 AGTGGCTATGTGGCTCCCGAGGCCCTAATTGGGAACAGTTTTCCTGATGCTTTGAGA 238
280 AGTACTTCTTTTACAGAAATTTTCAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 339
239 AGTACTTCTTTTACAGAAATTTTCAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297
340 GGAGCCATGGAATTTCTTCCCTTTTGGGAATATGTCCTTACCATTTTACATCTTTT 399
298 GGAGCCATGGAATTTCTTCCCTTTTGGGAATATGTCCTTACCATTTTACATCTTTT 357
400 GATGTCGACTAATGCCATTTATGTTGTTAATGCCGTTATTGG 438
358 GATGTCGACTAATGCCATTTATGTTGTTAATGCCGTTATTGG 396

RESULT 15
AA488304 376 bp mRNA linear EST 05-MAR-1998
DEFINITION re30a04.t1 Gessler Wilms tumor Homo sapiens cDNA clone IMAGE:897294
LOCUS AA488304
ACCESSION AA488304
(HUMAN); mRNA sequence.

VERSION AA488304.1 GI:2215735
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 376)

AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,
Kritzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin
J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theisling, B.,
White, T., Wyllie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

TITLE
JOURNAL
COMMENT

FEATURES
Source
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 1326 Std Error: 0.00
High quality sequence stop: 315.
Location/Qualifiers

1. 376

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:897294"

/clone_lib="Gessler Wilms tumor"

/sex="pooled (6)"

/lab_host="DH10b"

/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; RNA
was prepared from a pool of 6 anonymous Wilms' tumor RNAs.
RNA was prepared by acid-phenol, followed by one round of
oligo dt selection. cDNA library preparation was with
the BRL/Life Tech. Superscript plasmid system. An
oligo-dT NotI primer for first strand synthesis generated
ggggcgccg(c) at the 3' end of the clones. A 5' SalI
adapter was used with sequence 5'-gtcaccgacgctgcg-3'.
Resulting cDNAs were size selected (average size 2 kb).
NotI digested, and ligated into NotI/SalI-cut pSPORT1.
Library was constructed by Dr. Manfred Gessler."

BASE COUNT 97 a 77 c 75 g 127 t

ORIGIN

Query Match 24.8%; Score 330.4; DB 9; Length 376;
Best Local Similarity 95.5%; Pred. No. 2,9e-67;

Matches 340; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

370 AAATGCCCATTCAGCTTTTTCAGATCTTTGATGTCACATATGCAATTAATGTAATGC 429
|||||
1 AAATGCCCATTCAGCTTTTTCAGATCTTTGATGTCACATATGCAATTAATGTAATGC 60
430 CGTATGTTGATGTCACATATGTAATTAATTAATCTTACAGTAATCTTACACTTGATTT 489
|||||
61 CGTATGTTGATGTCACATATGTAATTAATTAATCTTACAGTAATCTTACACTTGATTT 120
490 GCTGCACTCTACCAATAGCTTTTGAATGACGTAAGTGTAAACAGAGAAAGAGCATG 549
|||||
121 GCTGCACTCTACCAATAGCTTTTGAATGACGTAAGTGTAAACAGAGAAAGAGCATG 180
550 TCTGAGAAAGAGTAAATATATTTTGGTACTTTATCTGAATCCCAAGATCTGCTT 609
|||||
181 TCTGAGAAAGAGTAAATATATTTTGGTACTTTATCTGAATCCCAAGATCTGCTT 240
610 CCCCTGCAAGTTGTTTCTTCTTACGATCTCATGTAATCCCTCTGGAGACACAGAC 669
|||||
241 CCCCTGCAAGTTGTTTCTTCTTACGATCTCATGTAATCCCTCTGGAGACACAGAC 300
670 AGTAGTAGAATCTCCATTTCTTCTTGTGTTTGAAGACAGAGACTGTCTC 725
|||||
301 AGTAGTAGAATCTCCATTTCTTCTTGTGTTTGAAGACAGAGACTGTCTC 356

Search completed: June 27, 2003, 16:18:06
Job time: 1831 secs